

GENERALIZED FLUCTUATION BOUNDS FOR STOCHASTIC ALGORITHMS IN THE PRESENCE OF COMPACTNESS

MORENIKEJI NERI^a, NICHOLAS PISCHKE^b, THOMAS POWELL^b

^a Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,

^b Department of Computer Science, University of Bath,
Claverton Down, Bath, BA2 7AY, United Kingdom,

E-mails: neri@mathematik.tu-darmstadt.de, nnp39@bath.ac.uk, trjp20@bath.ac.uk

ABSTRACT. We provide a convergence result for sequences of random variables taking values in a metric space that satisfy a stochastic quasi-Fejér monotonicity condition, in the context of a (local) compactness assumption. Our result is quantitative in that we derive an explicit and effective construction which, in terms of only a few moduli representing quantitative witnesses to key properties of the sequence of random variables and the underlying metric space involved, provides a metastable rate of pointwise convergence, a type of generalized fluctuation bound. That quantitative result in particular relies on the development of a finitary theory of martingales, culminating in a fully finitary Robbins-Siegmund theorem. We outline how this result particularises to the circumstances of the seminal work of Combettes and Pesquet on stochastic quasi-Fejér monotone sequences in separable Hilbert spaces, and we provide an initial application by illustrating how these results can be used to provide a metastable rate of pointwise convergence for a stochastic Krasnoselskii-Mann scheme solving a stochastic common fixed point problem for nonexpansive maps over proper Hadamard spaces. This work is set in the context of recent applications of the logic-based methodology of proof mining to probability theory, and represents its most sophisticated case study to date.

Keywords: stochastic quasi-Fejér monotonicity; martingale theory; Robbins-Siegmund theorem; generalized fluctuations; proof mining

MSC2020 Classification: 62L20, 90C15, 47J25, 60G42, 03F10

1. INTRODUCTION

1.1. Background and motivation. Fejér monotonicity is one of the most fundamental concepts in the modern study of numerical algorithms. At its core, it captures the idea that a method (x_n) in some metric space (X, d) decreases towards a target set $F \subseteq X$, or more formally:

$$\forall n \in \mathbb{N} \forall z \in F (d(x_{n+1}, z) \leq d(x_n, z)).$$

Extensions of this basic property have proven to be powerful organisational principles in the convergence analysis of algorithms, in which F generally represents some set of solutions to a problem under consideration. This is primarily because Fejér monotonicity occupies a unique position where (1) large classes of complex algorithms conform to one of its many variants and (2) it allows for a streamlined convergence analysis of associated methods, where additional properties which guarantee convergence can be articulated in a similarly abstract and uniform way.

Historically, the name Fejér monotonicity for the above concept was coined by Motzkin and Schoenberg [41], after a use made of a similar concept by Fejér in [27] as well as some subsequent work. While these works were based on Euclidean spaces, extensions to Hilbert spaces along with the development of a large amount of the related theory is due to Eremin

[22, 23] and Ermol'ev [24, 25, 26], together with other exponents of the “Russian school” of convex programming. Much of the modern appreciation of Fejér monotonicity as a unifying tool can be attributed to the work of Bauschke and Combettes, especially the surveys and expositions of Combettes [14, 15] and the central monograph [6]. Today, the zoo of variants of Fejér monotonicity is vast, incorporating, for example, varying metrics [19], adaptations to different distance functions like Bregman distances [5] or very general classes of distances on metric spaces [49] (see also [44]), or extensions and variations for different classes of spaces like Hadamard spaces [8] or general metric spaces [35, 36]. The most recent work in this direction includes so-called Fejér* monotonicity [2, 9, 10] as well as the extensive generalization of Fejér monotonicity introduced by Kohlenbach and Pinto [38]. Several of the aforementioned works also consider quantitative aspects in one way or the other.

In the case of stochastic algorithms, Fejér monotonicity naturally lifts to a corresponding supermartingale property, which when extended with error terms becomes satisfied by a vast array of stochastic methods. So-called stochastic quasi-Fejér monotonicity already originated in the work of Ermol'ev [24, 25, 26], and has more recently been isolated as an abstract concept in the more general setting of Hilbert spaces in the work of Combettes and Pesquet [17] (see also [18] for a quantitative perspective and [50] for extensions of the main abstract result from [17] to Hadamard spaces), where it is subsequently used to establish convergence of a range of optimization algorithms. In the generalised setting of [17], a stochastic algorithm (x_n) adapted to some filtration (\mathcal{F}_n) over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in some separable Hilbert space $(X, \|\cdot\|)$ is stochastically quasi-Fejér monotone w.r.t. a set $F \subseteq X$ if

$$(*) \quad \mathbb{E}[\phi(\|x_{n+1} - z\|) \mid \mathcal{F}_n] + \theta_n(z) \leq (1 + \chi_n(z))\phi(\|x_n - z\|) + \eta_n(z) \quad \mathbb{P}\text{-a.s.}$$

for all $z \in F$ and $n \in \mathbb{N}$, where $\phi : [0, \infty) \rightarrow [0, \infty)$ is strictly increasing and coercive (i.e. $\lim_{t \rightarrow +\infty} \phi(t) = +\infty$), and $(\theta_n(z))$ as well as $(\chi_n(z))$, $(\eta_n(z))$ are sequences of random variables adapted to (\mathcal{F}_n) , with the latter two assumed to be summable almost surely.

In the special case that $F = \{z\}$, i.e. in case the solution to the problem is actually unique, the convergence analysis of (stochastically) quasi-Fejér monotone sequences simplifies considerably, and it is often possible to prove strong convergence under minimal additional assumptions. Moreover, in such a case one can often provide explicit rates of convergence towards this solution (provided one has a sufficient quantitative rendering of the above uniqueness property). In the stochastic setting, a classic example of this is given by stochastic gradient methods for strongly convex functions (or more generally the Robbins-Monro procedure for strongly monotone operators), whose standard convergence proof via martingales fits into the above framework exactly. Building on preceding work for deterministic sequences in [36], such reasoning was exploited recently in work of the authors [44] to provide rates of convergence for stochastically quasi-Fejér monotone sequences in the context of uniqueness and in an abstract setting, in particular working over metric spaces and general classes of distance.¹ Similar results, albeit restricted to Hilbert spaces and specific forms of problems and types of regularity, can also be found in [18].

However, where solutions are not necessarily unique (or at least regular), establishing strong convergence is more subtle, and generally relies on a (local) compactness assumption. Crucially, and already for deterministic algorithms, effective rates of convergence are then generally no

¹Such results can even be generalized towards a selected collection of non-unique problems, provided the problems are sufficiently regular. An abstract study of such suitable regularity conditions was given in [36] for deterministic sequences. An adaptation of such general results to the case of stochastic sequences is forthcoming work by the second and third author.

longer possible.² The resulting problem of establishing suitable quantitative results for deterministic Fejér monotone sequences in the presence of compactness in such general cases was first considered by Kohlenbach, Leuştean and Nicolae [35]. Roughly speaking, there it is shown that if (x_n) is a quasi-Fejér monotone sequence in some compact space X with respect to some F that is closed in a certain explicit sense, and (x_n) satisfies an approximation property with respect to F , then (x_n) strongly converges to a point in F , and moreover, given suitable quantitative versions of these assumptions, an explicit and highly uniform bound for $\exists N$ in the following *metastable* formulation of the Cauchy property can be given:

$$\forall k \in \mathbb{N} \forall g : \mathbb{N} \rightarrow \mathbb{N} \exists N \in \mathbb{N} \forall i, j \in [N; N + g(N)] \left(d(x_i, x_j) \leq \frac{1}{k+1} \right),$$

where we write $[n; m] := [n, m] \cap \mathbb{N}$. Metastable convergence (and more generally metastable versions of infinitary concepts) has been popularised by Tao (see [55] and also [54]), and are by now frequently utilised across various parts of analysis and optimization as it is typically possible to find explicit rates of metastability (i.e. bounds on $\exists N$ above, in terms of k and g) even in situations where computable rates of convergence or the like do not exist, and thus metastability provides a much needed substitute for quantitative information on the convergence behaviour in these cases. Informally speaking, rates of metastability represent an alternative quantitative characterisation of convergence in terms of generalised *fluctuation* behaviour of the sequence (x_n) , providing for arbitrary functions $g : \mathbb{N} \rightarrow \mathbb{N}$ a bound on the existence of an interval $[N; N + g(N)]$ in which no $1/(k+1)$ -fluctuations can occur.³

These first general quantitative results for Fejér monotone sequences in the presence of compactness date back to almost a decade ago, where they are roughly contemporaneous with the revived study of stochastic quasi-Fejér monotonicity for non-unique solution sets mentioned previously. However, quantitative results for the latter were never considered, and indeed the underlying techniques from mathematical logic that facilitated the construction of rates of metastability in [35] (see the discussion in the following sections, in particular Section 1.3) were, with a few isolated exceptions, never deployed in a probabilistic setting until a few years ago. Here one is confronted with a series of significant obstacles, including the increased logical complexity of almost-sure statements, and the arsenal of highly technical machinery required for the convergence analysis of stochastic quasi-Fejér monotone sequences, chief among them the Robbins-Siegmund supermartingale convergence theorem [52].

Leveraging some recent results which bring these logical techniques to bear on probability for the first time, we tackle these obstacles and provide a first quantitative study of stochastic algorithms in the presence of compactness, representing simultaneously a first quantitative analysis of stochastic quasi-Fejér monotonicity for non-unique solution sets, and a stochastic extension of [35] of considerable additional complexity. Our main achievements include new, general notions of stochastic quasi-Fejér monotonicity, convergence theorems that are both qualitatively and quantitatively novel, and in particular come equipped with fluctuation information, and the development of a fully finitary theory of supermartingales that is both necessary for our main results and represents a base for future work in quantitative stochastic optimization. The following section now discusses these contributions at some further depth.

1.2. Overview of the main technical results. Our central object of study is a new and highly general finitary variant of stochastic quasi-Fejér monotonicity for algorithms taking

²This can in fact be made formally precise using tools from mathematical logic and computability theory, using so-called Specker sequences [53] (see in particular also [48]). We refer to [35, 36] for further discussions.

³For more details on how rates of metastability can be viewed as generalised fluctuation bounds see [39], and for stochastic processes the more recent [47].

values in a general metric space (X, d) . Taking inspiration from [35], we consider solution sets that can be decomposed in an abstract way as $F = \bigcap_{k \in \mathbb{N}} AF_k$ via a sequence of measurable approximation sets $AF_0 \subseteq AF_1 \subseteq \dots$, giving us the flexibility to uniformly capture various problem formulations and associated notions of approximate solutions from the literature on stochastic optimization.

A sequence (x_n) adapted to some filtration (\mathcal{F}_n) as above is then called *uniformly stochastically quasi-Fejér monotone* w.r.t. F , (AF_k) and (\mathcal{F}_n) if⁴

$$\forall \lambda > 0 \forall r, n \in \mathbb{N} \exists k \in \mathbb{N} \forall z \in AF_k \forall m \leq n \left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1} \right) < \lambda \right),$$

where $G : [0, \infty) \rightarrow [0, \infty)$ is a strictly increasing, coercive and continuous function with $G(0) = 0$,⁵ and $(\eta_m), (\chi_n)$ are adapted to (\mathcal{F}_n) and almost surely summable.⁶ While discussed later in more detail, a rather direct compactness argument in fact shows that over finite-dimensional Hilbert spaces and under suitable boundedness assumptions on the solution set F and its approximations AF_k , this property is implied by the previously considered notion $(*)$ (for error terms uniform in z).

The main result of the paper is a convergence theorem for such uniformly stochastically quasi-Fejér monotone sequences, in which almost sure convergence of (x_n) towards some F -valued random variable x is established under a (local) compactness assumption on the space, a closure assumption of F , and a rather weak stochastic approximation property on the sequence (x_n) formulated using (AF_k) (cf. Theorems 4.21 and 4.22). These theorems, while novel in their own right, are however merely qualitative renderings of a full quantitative convergence theorem (cf. Theorem 4.18) which provides explicit generalized fluctuation information on the sequence (x_n) in the form of a bound $\Delta(\lambda, k, g)$ satisfying

$$\mathbb{P} \left(\forall n \leq \Delta(\lambda, k, g) \exists i, j \in [n; n + g(n)] \left(d(x_i, x_j) > \frac{1}{k+1} \right) \right) < \lambda,$$

for all $\lambda > 0$, $k \in \mathbb{N}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$, a so-called metastable rate of pointwise convergence for stochastic quasi-Fejér monotone sequences (also known as a λ -uniform bound for the ε -metastable pointwise convergence in [4], where the notion was first isolated in a logical context). The bound Δ can be explicitly constructed from some simple data representing quantitative formulations of the main assumptions, making it highly uniform and in particular independent of the distribution of the underlying probability space.

We envisage our quantitative theorem being used in one of two ways: In straightforward scenarios the rather intricate general construction will simplify considerably, and then, potentially with the help of ad-hoc optimizations, could be used to provide low-complexity certificates on the fluctuation behaviour of the algorithm in question. Alternatively, where the exact definition

⁴This notion omits the negative correction term (θ_n) from $(*)$, which is generally superfluous for a quasi-Fejér type argument and often used simply for establishing an associated approximation property, which we deal with separately in this paper.

⁵Compared to the ϕ used in $(*)$, which is only continuous almost everywhere, we here assume that the function G is continuous everywhere and also satisfies $G(0) = 0$, as this later allows for an easier handling of approximation errors which is crucial for our arguments. However, as most functions employed in the context of Fejér monotonicity satisfy these extended assumptions, this is a very mild restriction.

⁶The move to an approximate solution $z \in AF_k$ in the above notion generally renders the dependence of $(\eta_m), (\chi_n)$ on z occurring in $(*)$ senseless, which is why our notion considers the errors to be uniformly chosen. Practically, this not a great sacrifice as many, if not most, methods can be treated in the context of $(*)$ already without such a dependence on z .

of Δ is of less interest, its uniformity properties can still be invoked to characterise the data on which such generalized fluctuation bounds depend.

A concrete example capturing all of this is provided in Section 5. Here we consider the case where F is the solution set of a stochastic common fixed-point problem, i.e.

$$F := \{z \in X \mid T_v z = z \text{ } \mathbb{P}\text{-a.s.}\},$$

where $(T_v)_{v \in V}$ is a suitably measurable family of nonexpansive mappings and v is a V -valued random variable, for a measurable space (V, \mathcal{V}) . Moreover, X in this case is taken to be a separable Hadamard spaces, that is a complete separable metric spaces of nonpositive curvature in the sense of Alexandrov, making use of our general metric setting. In that context, our main results are then used to prove the convergence of a stochastic Krasnoselskii-Mann scheme⁷

$$x_{n+1} := (1 - \lambda_n)x_n \oplus \lambda_n T_{v_n} x_n$$

under standard conditions on the step-sizes (λ_n) and the samples (v_n) of v (see Theorem 5.12), where the bound Δ is used in the second way as indicated above merely to highlight uniformities of the fluctuation information. We envisage that many more sophisticated examples of this kind are possible, and further applications will be provided in future work.

Establishing our main quantitative result requires all of the machinery that traditionally underlies the convergence of stochastic algorithms to be reformulated in a fully finitary manner, resolving both (in)equalities and probabilities via approximate errors. This includes Doob's martingale convergence theorem [20] and its well-known extension to almost-supermartingales by Robbins and Siegmund [52]. In that way, the entire first half of the paper is dedicated to the development of a fully finitary supermartingale convergence theory. For that, we build on initial work in this direction by the first and third author [45, 46] which provides basic fluctuation information for (almost) supermartingales, but here we require a much more careful quantitative analysis, where our definition of stochastic quasi-Fejér monotonicity requires us to provide fluctuation information for a weaker kind of process satisfying only an approximate finitary variant of the martingale property, namely so-called λ - ε - N -supermartingales, where the usual property

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n \text{ } \mathbb{P}\text{-a.s., for all } n \in \mathbb{N}$$

is replaced by the weaker property

$$\mathbb{P}(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] > X_n + \varepsilon) < \lambda, \text{ for all } n \leq N.$$

The resulting quantitative supermartingale convergence theorem is captured as Theorem 2.15, and along the way involves finitary analogues of the Doob decomposition theorem (Lemma 2.3), the optional stopping theorem (Lemma 2.9), and crossing inequalities (Lemma 2.14), along with a subtle quantitative treatment of integrability via so-called moduli of absolute continuity (as introduced in [51]). This is then extended in a similar way to the full Robbins-Siegmund theorem (Theorem 3.9). All of these results are entirely independent of the subsequent analysis of Fejér monotone sequences, and can potentially be exploited in any situation where a finitary analogue of martingale convergence is required. Even beyond this, on our route to the main theorem in Section 4, we introduce and study a number of general notions which we conjecture are of use beyond the goals of this paper. These include results on stochastic lim inf-moduli, and particularly the introduction of the concept of *set-fluctuations* (see Lemma 4.14). We therefore consider the development of this machinery to be a contribution in its own right, independently of our main result on Fejér monotonicity.

⁷Here, given $x, y \in X$ and $\lambda \in [0, 1]$, we write $(1 - \lambda)x \oplus \lambda y$ for the unique point z on the geodesic connecting x and y satisfying $d(x, z) = \lambda d(x, y)$ and $d(y, z) = (1 - \lambda)d(x, y)$, as it exists in Hadamard spaces.

1.3. Relations to logic and proof mining. We intend our results to be primarily of practical use within stochastic optimization, or in the case of our finitary martingale convergence theory, of potentially broader relevance within quantitative probability theory. However, as already hinted at before, methodologically this paper belongs to the logic-based proof mining program, and it takes on an additional significance when viewed from this perspective. Proof mining is a subfield of mathematical logic which aims at repurposing tools and techniques from the foundations of mathematics to provide additional computational information for ineffectively proven theorems in mathematics. As such, it has found wide success in particular in the context of (nonlinear) analysis and optimization, and we refer to the seminal monograph [32] as well as the surveys [33, 37] for further background on both the theory and practice of this field.

The extent to which logical insights and tools play a role in each of the many quantitative studies that are rooted in proof mining varies. In some cases, logic is used with a very light touch, often manifesting only in a certain abstract approach, or as a guide for the author in choosing the correct moduli to represent key assumptions in a quantitative way. This paper, on the other hand, makes extensive use of the logical machinery, and in terms of its overall scope and the complexity of the quantitative analysis, represents one of the most logically intricate case studies to come out of the proof mining tradition, in particular illustrated by the complexity of the general construction of the stochastic metastable bound Δ .

Indeed, this paper forms a synthesis of several different strands of research in proof mining, in particular its most recent and advanced developments. First and foremost, we build on a series of papers giving quantitative meaning to Fejér monotone sequences in the nonstochastic setting, starting with [35] and continuing in [38, 49], among others, similarly originating from this logical point of view. This then is lifted to stochastic sequences by further developing several recent works on quantitative aspects of martingales [45, 46], which in turn build on older results in ergodic theory [4], all likewise applications of proof mining. Finally, we make heavy use of logical insights codified by the first and second author in their so-called logical metatheorems for probability theory [43], the logical foundations for applications of proof mining to that area, and further refined in the forthcoming work by the first and second author with Oliva [42], particularly for arriving at the correct resolution of almost-sure statements. Lastly, the present paper complements and extends previous work of the authors on the quantitative convergence behavior of stochastic quasi-Fejér monotone sequences under uniqueness assumptions [44]. As such, this paper represents a tour through many of the key recent developments in the proof mining programme, all of which are brought together to solve the problem at hand: Making quantitative sense of stochastic quasi-Fejér monotonicity in the presence of compactness.

Outside of proof mining it is hoped that our results will be of interest to a general logical audience as a particularly striking illustration of Hilbert’s programme in action. We analyse a thoroughly modern convergence framework that makes use of a substantial theory which at first glance carries considerable set-theoretic complexity, encompassing probability measures, random variables (in a metric setting), Lebesgue integrals, conditional expectations, martingale convergence, stopping times, compactness arguments, and much more. We are nevertheless able to reduce all of this to a simple finitary core in the form of Theorem 4.18. That a proof of this nature is amenable to such a treatment is further demonstration of the phenomenon of ‘proof-theoretic tameness’ (see [34]), and another vindication of Hilbert’s programme as applied to mainstream mathematics.

All these logical aspects notwithstanding, everything that follows is formulated without explicit reference to logic and can be comprehended by a reader with no knowledge of proof mining, a typical feature of applications of proof mining. To the reader interested in some of

the underlying logical aspects, we include a few brief remarks highlighting some of the logical aspects.

1.4. Summary and organization of the paper. In summary, our main contributions include the following, each of which may be of interest to a slightly different audience:

- A new, general notion of stochastic quasi-Fejér monotonicity in a metric setting (Definition 4.6), and a novel strong convergence result utilising a compactness argument (Theorems 4.21 and 4.22).
- A quantitative rendering of that convergence result, providing highly uniform generalized fluctuation information for the algorithm in the form of a stochastic metastable bound (Theorem 4.18).
- A concrete case study establishing the convergence of a randomized Krasnoselskii-Mann scheme for approximating common stochastic fixed points of nonexpansive mappings in Hadamard spaces (Theorem 5.12).
- The development of a finitary theory for supermartingale convergence (Section 2) and its extension to almost-supermartingales (Section 3), culminating in a fully finitized Robbins-Siegmund theorem (Theorem 3.9).
- An intricate demonstration of proof mining utilising a wide selection of recent advances in that area, and, from a more philosophical perspective, an illustration of how Hilbert's programme extends to modern probability theory and stochastic programming.

Indeed, the paper is organised so that it can be read in a different way by people with different background. For example, Sections 2 and 3 can be treated as a shorter standalone article on the finitary development of the Robbins-Siegmund theorem. Our quantitative framework for stochastic quasi-Fejér monotone sequences in Section 4 only requires the quantitative Robbins-Siegmund theorem at exactly one point and is otherwise entirely independent of the first part of the paper. Finally, all results can be understood in a purely qualitative way and used without reference to the finitary results as illustrated by the applications in Section 5.

1.5. Preliminaries and notation. Throughout this paper, we fix an ambient probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If not specified otherwise, all probabilistic notions such as measurability, random variables, almost sureness, etc., are understood relative to that space. All properties as well as (in)equalities between random variables are understood to hold only almost surely, if not stated otherwise. We denote the indicator function of a measurable set $A \in \mathcal{F}$ by $\mathbf{1}_A$. We will later be concerned with variables taking values in metric spaces (X, d) , in which case we understand the notion of measurability or random variable to be defined using the Borel σ -algebra $\mathcal{B}(X)$ of that space.

Further, we fix a filtration $\mathcal{F} := (\mathcal{F}_n)$ on that space. Given that filtration, we write $l_+(\mathcal{F})$ for the set of sequences of non-negative random variables (ξ_n) that are adapted to the filtration, i.e. where ξ_n is \mathcal{F}_n -measurable for all $n \in \mathbb{N}$. Further, we write $l_+^1(\mathcal{F})$ for the set of all $(\xi_n) \in l_+(\mathcal{F})$ such that $\sum_{n=0}^{\infty} \xi_n < +\infty$. We also call a sequence of random variables (x_n) a stochastic process, and we say that this process is adapted to \mathcal{F} if x_n is \mathcal{F}_n -measurable for all $n \in \mathbb{N}$.

To specify complex events, we follow the slightly nonstandard approach of using quantifiers instead of unions and intersections. Concretely, let $\varphi(n) \in \mathcal{F}$ be a measurable property for any $n \in \mathbb{N}$. Then we write

$$\exists n \varphi(n) := \bigcup_{n \in \mathbb{N}} \varphi(n) \text{ and } \forall n \varphi(n) := \bigcap_{n \in \mathbb{N}} \varphi(n).$$

Naturally, this extends to properties with an arbitrary finite amount of variables and any type of quantifier prefix. Crucially, we can use the continuity of the probability measure to derive the following type of “prenexation” results for the above notation:

Lemma 1.1 (Lemma 3.1 in [45]). *Let $\varphi(n) \in \mathbf{F}$ be a measurable property for all $n \in \mathbb{N}$ and $p \in [0, 1]$. If φ is anti-monotone, i.e. $\varphi(n+1) \subseteq \varphi(n)$ for all $n \in \mathbb{N}$, then*

- (1) $\mathbb{P}(\forall n \varphi(n)) \geq p$ if, and only if, $\forall n (\mathbb{P}(\varphi(n)) \geq p)$,
- (2) $\mathbb{P}(\forall n \varphi(n)) \leq p$ if, and only if, $\forall \lambda > 0 \exists n (\mathbb{P}(\varphi(n)) < p + \lambda)$.

If φ is monotone, i.e. $\varphi(n) \subseteq \varphi(n+1)$ for all $n \in \mathbb{N}$, then

- (1) $\mathbb{P}(\exists n \varphi(n)) \leq p$ if, and only if, $\forall n (\mathbb{P}(\varphi(n)) \leq p)$,
- (2) $\mathbb{P}(\exists n \varphi(n)) \geq p$ if, and only if, $\forall \lambda > 0 \exists n (\mathbb{P}(\varphi(n)) > p - \lambda)$.

Lastly, to talk about intervals of natural numbers, we use the notation $[n; m] := [n, m] \cap \mathbb{N}$ throughout.

2. FINITARY MARTINGALES AND QUANTITATIVE MARTINGALE CONVERGENCE

In order to give a quantitative treatment to stochastic quasi-Fejér monotone sequences, we first need to develop a quantitative theory of martingales. This task will occupy us for this as well as the following section and culminates in an quantitative variant of the Robbins-Siegmund theorem on almost-supermartingales [52]. As such, our whole approach relies on, and extends, existing work on quantitative aspects of martingale convergence, as developed recently by the first and last author [45, 46], in two crucial ways:

First, supermartingales are replaced throughout with the weaker, finitary notion of being a λ - ε - N -supermartingale, a quantitative variant of the usual notion of a supermartingale which

- (1) concerns only finite fragments of the process with length $N \in \mathbb{N}$,
- (2) relaxes the descent condition by including an approximate error $\varepsilon > 0$,
- (3) weakens the almost-sure to being false with low probability $\lambda > 0$.

Switching to this variant requires a careful adaptation of several standard results on martingales to the finitary setting, along with a corresponding strengthening of the recent quantitative martingale convergence theorem [45, Theorem 7.4], which did not feature such considerations as there the supermartingale property was not resolved.⁸

Second, we use this new quantitative martingale convergence theorem to provide an alternative quantitative variant of the Robbins-Siegmund theorem compared to that recently given in [46] by not only strengthening the result to finitary almost supermartingales, similar to above, but with convergence now formulated in terms of *fluctuations*, a crucial adjustment for the delicate calculations around Fejér monotone sequences that follow.

The main notion that we henceforth work with is the following finitary weakening of a supermartingale:

Definition 2.1 (Finitary supermartingale). Let $\lambda, \varepsilon > 0$ and let (X_n) be a sequence of random variables adapted to (\mathcal{F}_n) with $\mathbb{E}[|X_n|] < +\infty$ for all $n \in \mathbb{N}$. The process (X_n) is called a λ - ε -supermartingale if

$$\mathbb{P}(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] > X_n + \varepsilon) < \lambda$$

for all $n \in \mathbb{N}$. It is called a λ - ε - N -supermartingale if this property is only required to hold for all $n < N$.

⁸In logical terms, it treated the associated theorems on supermartingales in their rule form.

Analogously, we call (X_n) a λ - ε - $(-N)$ -submartingale if the displayed property is replaced by the requirement that

$$\mathbb{P}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] < X_n - \varepsilon) < \lambda$$

for any $n \in \mathbb{N}$ (any $n < N$). Clearly, (X_n) is a λ - ε - $(-N)$ -supermartingale if, and only if, $(-X_n)$ is a λ - ε - $(-N)$ -submartingale.

Remark 2.2 (For logicians). The above property of a λ - ε - N -supermartingale is a full finitization of the usual supermartingale property, following the works on the logical foundations of proof mining and probability [42, 43], an aspect that crucially underlies the conception of this paper. Concretely, these logical results guarantee that there are suitable quantitative variants based on finitary supermartingales for large classes of theorems phrased for ordinary supermartingales. The following quantitative results on supermartingales all arose through such a logical analysis of their respective infinitary counterparts.

The main result of the present section is a quantitative variant of the martingale convergence theorem phrased for λ - ε - N -supermartingales. For that, as mentioned above, we need to extend a range of standard results for ordinary supermartingales (an account of which can be found in any standard textbook in advanced probability, see e.g. [31]) to this more general setting of finitary supermartingales.

The first such result is a quantitative version of Doob's decomposition theorem, which allows one to decompose a stochastic process (X_n) into a sum $X_n = Y_n + Z_n$ such that (Y_n) is a martingale and (Z_n) a predictable process (i.e. Z_n is \mathcal{F}_{n-1} measurable for all $n \geq 1$), which is moreover nonincreasing whenever (X_n) is a supermartingale.

Lemma 2.3. *Let (X_n) be adapted to (\mathcal{F}_n) and be such that $\mathbb{E}[|X_n|] < +\infty$ for all $n \in \mathbb{N}$. Then $X_n = Y_n + Z_n$ for any $n \in \mathbb{N}$ where (Y_n) is a martingale w.r.t. (\mathcal{F}_n) and (Z_n) is predictable w.r.t. (\mathcal{F}_n) . Furthermore, if (X_n) is a λ - ε - N -supermartingale, then (Z_n) is λ - ε - N -nonincreasing, i.e.*

$$\mathbb{P}(Z_{n+1} > Z_n + \varepsilon) < \lambda$$

for all $n < N$.

Proof. We proceed as usual in the proof of Doob's decomposition theorem and define $Z_0 := 0$ as well as

$$Z_n := \sum_{k=1}^n (\mathbb{E}[X_k | \mathcal{F}_{k-1}] - X_{k-1})$$

for $n \geq 1$. Then each Z_n is \mathcal{F}_{n-1} -measurable and (Z_n) is hence predictable. We further define $Y_n := X_n - Z_n$ and it follows as usual that $\mathbb{E}[|Y_n|] < +\infty$ and that (Y_n) is a martingale w.r.t. (\mathcal{F}_n) . If (X_n) is now a λ - ε - N -supermartingale, i.e. we have

$$\mathbb{P}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] > X_n + \varepsilon) < \lambda$$

for any $n < N$, then for any $\omega \in \Omega$ with $\mathbb{E}[X_{n+1} | \mathcal{F}_n](\omega) \leq X_n(\omega) + \varepsilon$, we have that

$$\begin{aligned} Z_{n+1}(\omega) &= \sum_{k=1}^{n+1} (\mathbb{E}[X_k | \mathcal{F}_{k-1}](\omega) - X_{k-1}(\omega)) \\ &= Z_n(\omega) + (\mathbb{E}[X_{n+1} | \mathcal{F}_n](\omega) - X_n(\omega)) \\ &\leq Z_n(\omega) + \varepsilon. \end{aligned}$$

Thus we have shown that

$$\mathbb{P}(Z_{n+1} > Z_n + \varepsilon) \leq \mathbb{P}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] > X_n + \varepsilon) < \lambda$$

for any $n < N$, i.e. that (Z_n) is λ - ε - N -nonincreasing. \square

The next result we consider is Doob's optional stopping theorem, that is that for bounded stopping times $\rho \leq \tau$, we get $\mathbb{E}[X_\tau \mid \mathcal{F}_\rho] \leq X_\rho$ for a supermartingale (X_n) and the stopped filtration defined via

$$\mathcal{F}_\rho := \{A \in \mathcal{F} \mid A \cap \{\rho \leq t\} \in \mathcal{F}_t \text{ for all } t \in \mathbb{N}\}.$$

Before we can give a quantitative version of this result, we need to develop a few further (quantitative) properties of random variables and martingales.

The first such result is quantitative version of the fact that the expected value of an ordinary supermartingale decreases. This property transfers to an approximate version for finitary supermartingales. However, to give such a quantitative result, we have to quantitatively resolve the property that each respective random variable is integrable in a more comprehensive way than simply providing an upper bound on the mean.

The enriched data we rely on here is a so-called modulus of absolute continuity:

Definition 2.4. Let X be an integrable random variable. We call a function $\mu : (0, \infty) \rightarrow (0, \infty)$ such that

$$\forall \varepsilon > 0 \forall A \in \mathcal{F} (\mathbb{P}(A) < \mu(\varepsilon) \rightarrow \mathbb{E}[|X|1_A] \leq \varepsilon)$$

a modulus of absolute continuity for X .

By e.g. Lemma 13.1 in [57], such a modulus always exists for integrable X . However, the quantitative information it provides complements that of a simple upper bound on the mean, with both together providing a pair of data fully characterising the integrability of a random variable quantitatively. Later, in Section 5, we discuss concrete instantiations for such moduli in the context of applications to stochastic approximation methods.

Remark 2.5 (For logicians). The above notion of a modulus of absolute continuity is motivated by the proof-theoretic methodology of proof mining. In particular, the underlying logical results (see again [43] and in particular also [42]) guarantee that such moduli can be extracted from large classes of proofs of the integrability of a random variable.

Using this modulus, we can then immediately obtain the following quantitative descent property for finitary martingales.

Lemma 2.6. Let μ_n be a modulus of absolute continuity for X_n . Further, let (X_n) be a $\mu_N^M(\varepsilon/4N)$ - $\varepsilon/2N$ - N -supermartingale, where $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$. Then

$$\mathbb{E}[X_m] \leq \mathbb{E}[X_n] + \varepsilon$$

for all $n \leq m \leq N$.

Proof. As (X_n) is a $\mu_N^M(\varepsilon/4N)$ - $\varepsilon/2N$ - N -supermartingale, we get

$$\mathbb{P}(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] > X_n + \varepsilon/2N) < \mu_N^M(\varepsilon/4N) \leq \mu_{n+1}(\varepsilon/4N), \mu_n(\varepsilon/4N)$$

for any $n < N$. Denote the set inside the probability above by A_n . Then we have

$$\begin{aligned} \mathbb{E}[X_{n+1}] &= \mathbb{E}[X_n] + \mathbb{E}[(X_{n+1} - X_n)\mathbf{1}_{A_n}] + \mathbb{E}[(X_{n+1} - X_n)\mathbf{1}_{A_n^c}] \\ &\leq \mathbb{E}[X_n] + \mathbb{E}[|X_{n+1}| \mathbf{1}_{A_n}] + \mathbb{E}[|X_n| \mathbf{1}_{A_n}] + \mathbb{E}[(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n)\mathbf{1}_{A_n^c}] \\ &\leq \mathbb{E}[X_n] + \varepsilon/4N + \varepsilon/4N + \varepsilon/2N \\ &= \mathbb{E}[X_n] + \varepsilon/N \end{aligned}$$

where the second line follows since $A_n^c \in \mathcal{F}_n$, and the third using $\mathbb{P}(A_n) < \mu_{n+1}(\varepsilon/4N), \mu_n(\varepsilon/4N)$ and that $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n \leq \varepsilon/2N$ holds on A_n^c . This in particular yields

$$\mathbb{E}[X_m] \leq \mathbb{E}[X_n] + (m - n)\varepsilon/N \leq \mathbb{E}[X_n] + \varepsilon$$

for all $n \leq m \leq N$. \square

Before moving on, we also require a quantitative version of the property that $\mathbb{E}[X \mid \mathbf{F}_0] \leq 0$ if, and only if, $\mathbb{E}[X\mathbf{1}_A] \leq 0$ for all $A \in \mathbf{F}_0$, whenever X is an integrable random variable and \mathbf{F}_0 is a given sub- σ -algebra of \mathbf{F} .

Lemma 2.7. *Let X be an integrable random variable and \mathbf{F}_0 be a given sub- σ -algebra of \mathbf{F} .*

- (1) *For any $\varepsilon, \lambda > 0$, if $\mathbb{E}[X\mathbf{1}_A] < \varepsilon\lambda$ for all $A \in \mathbf{F}_0$, then $\mathbb{P}(\mathbb{E}[X \mid \mathbf{F}_0] > \varepsilon) < \lambda$.*
- (2) *Let μ be a modulus of absolute continuity for X . For any $\varepsilon > 0$, if $\mathbb{P}(\mathbb{E}[X \mid \mathbf{F}_0] > \varepsilon/2) < \mu(\varepsilon/2)$, then $\mathbb{E}[X\mathbf{1}_A] < \varepsilon$ for all $A \in \mathbf{F}_0$.*

Proof. (1) Suppose for a contradiction that $\mathbb{P}(\mathbb{E}[X \mid \mathbf{F}_0] > \varepsilon) \geq \lambda$, and write A_0 for the set inside the probability. Note that $A_0 \in \mathbf{F}_0$. But then it holds that

$$\varepsilon\lambda > \mathbb{E}[X\mathbf{1}_{A_0}] = \mathbb{E}[\mathbb{E}[X \mid \mathbf{F}_0]\mathbf{1}_{A_0}] \geq \varepsilon\mathbb{P}(A_0) \geq \varepsilon\lambda,$$

which is a contradiction.

- (2) Assume $\mathbb{P}(\mathbb{E}[X \mid \mathbf{F}_0] > \varepsilon/2) < \mu(\varepsilon/2)$, and write A_0 for the set inside the probability. As before, we have $A_0 \in \mathbf{F}_0$, and so given an arbitrary $A \in \mathbf{F}_0$, we have $A \cap A_0 \in \mathbf{F}_0$. This yields

$$\begin{aligned} \mathbb{E}[X\mathbf{1}_A] &= \mathbb{E}[X\mathbf{1}_{A \cap A_0}] + \mathbb{E}[X\mathbf{1}_{A \cap A_0^c}] \\ &\leq \mathbb{E}[|X|\mathbf{1}_{A \cap A_0}] + \mathbb{E}[\mathbb{E}[X \mid \mathbf{F}_0]\mathbf{1}_{A \cap A_0^c}] \\ &\leq \varepsilon/2 + \varepsilon/2 = \varepsilon, \end{aligned}$$

where, in the second line we have used that $A \cap A_0^c \in \mathbf{F}_0$ and in the last line that $\mathbb{P}(A \cap A_0) < \mu(\varepsilon/2)$. \square

We now turn to stopping times. The first result we provide there is a quantitative version of the result that a boundedly stopped process is integrable. For simplicity, we restrict ourselves to nonnegative processes.

Lemma 2.8. *Let τ be a bounded stopping time with $\tau \leq N$. Further, let (X_n) be a nonnegative stochastic process, and let μ_n be a modulus of absolute continuity for each X_n . Then $\mu_N^M(\varepsilon/N) := \min\{\mu_n(\varepsilon/N) \mid n \leq N\}$ is a modulus of absolute continuity for X_τ .*

Proof. Note that $X_\tau \leq \max_{n \leq N} X_n$ almost surely, so that

$$\mathbb{E}[X_\tau\mathbf{1}_A] \leq \mathbb{E}\left[\max_{n \leq N}(X_n\mathbf{1}_A)\right] \leq \sum_{n \leq N} \mathbb{E}[X_n\mathbf{1}_A] \leq N(\varepsilon/N) = \varepsilon$$

for A such that $\mathbb{P}(A) < \mu_N^M(\varepsilon/N) \leq \mu_n(\varepsilon/N)$ for all $n \leq N$. \square

We are now in the position to provide a fully quantitative variant of Doob's optional stopping theorem, phrased for finitary martingales using moduli of absolute continuity. Again, we restrict ourselves to nonnegative processes.

Lemma 2.9. *Let ρ, τ be bounded stopping times w.r.t. (\mathcal{F}_n) with $\rho \leq \tau \leq N$. Further, let (X_n) be a nonnegative stochastic process, and let μ_n be a modulus of absolute continuity for each X_n . If (X_n) is a $\mu_N^M(\varepsilon\lambda/16N)$ - $(\varepsilon\lambda/8N)$ - N -supermartingale w.r.t. (\mathcal{F}_n) , then*

$$\mathbb{P}(\mathbb{E}[X_\tau \mid \mathcal{F}_\rho] > X_\rho + \varepsilon) < \lambda.$$

Here, as before: $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$.

Proof. Let A_0 denote the set of all ω such that $\rho(\omega) \leq \tau(\omega) \leq N$ as well as

$$X_{n+1}(\omega) \leq X_n(\omega) + (\varepsilon\lambda/8N)$$

for all $n < N$. Note that $\mathbb{P}(A_0^c) < \mu_N^M(\varepsilon\lambda/16N)$ by assumption on (X_n) . Then, for any $\omega \in A_0$, we have

$$\begin{aligned} X_{\tau(\omega)}(\omega) &= X_{\rho(\omega)}(\omega) + \sum_{n=\rho(\omega)}^{\tau(\omega)-1} (X_{n+1}(\omega) - X_n(\omega)) \\ &= X_{\rho(\omega)}(\omega) + \sum_{n=0}^{N-1} (X_{n+1}(\omega) - X_n(\omega)) \mathbf{1}_{\rho \leq n < \tau}(\omega). \end{aligned}$$

Now, for $A \in \mathcal{F}_\rho$, we have by the above that

$$\mathbb{E}[X_\tau \mathbf{1}_{A \cap A_0}] \leq \mathbb{E}[X_\rho \mathbf{1}_{A \cap A_0}] + \sum_{n=0}^{N-1} \mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_{A \cap A_0}].$$

Note that $\{\rho \leq n < \tau\} \cap A \in \mathcal{F}_n$ as $A \in \mathcal{F}_\rho$. Therefore, since $\mu_N^M(\varepsilon/2)$ is a modulus of absolute continuity for $X_{n+1} - X_n$, and since

$$\mathbb{P}(\mathbb{E}[X_{n+1} - X_n \mid \mathcal{F}_n] > \varepsilon\lambda/8N) < \mu_N^M(\varepsilon\lambda/16N),$$

part (2) of Lemma 2.7 yields that

$$\mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_A] < \varepsilon\lambda/4N$$

for all $n < N$. In particular, we have

$$\begin{aligned} \mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_{A \cap A_0}] &= \mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_A] - \mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_{A \cap A_0^c}] \\ &\leq \mathbb{E}[(X_{n+1} - X_n) \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_A] + \mathbb{E}[X_n \mathbf{1}_{\rho \leq n < \tau} \mathbf{1}_{A \cap A_0^c}] < \varepsilon\lambda/2N. \end{aligned}$$

We have thus shown that $\mathbb{E}[(X_\tau - X_\rho) \mathbf{1}_{A \cap A_0}] \leq \varepsilon\lambda/2$. By Lemma 2.8, $\mu_N^M(\varepsilon/N)$ is a modulus of absolute continuity for X_τ . Thereby, we have

$$\mathbb{E}[(X_\tau - X_\rho) \mathbf{1}_A] = \mathbb{E}[(X_\tau - X_\rho) \mathbf{1}_{A \cap A_0}] + \mathbb{E}[(X_\tau - X_\rho) \mathbf{1}_{A \cap A_0^c}] \leq \varepsilon\lambda/2 + \varepsilon\lambda/16 < \varepsilon\lambda$$

as $\mathbb{P}(A_0^c) < \mu_N^M(\varepsilon\lambda/16N)$. Part (1) of Lemma 2.7 now yields that

$$\mathbb{P}(\mathbb{E}[X_\tau \mid \mathcal{F}_\rho] > X_\rho + \varepsilon) < \lambda$$

as claimed. \square

Remark 2.10. If (X_n) is already a finitary martingale in the stronger sense that the associated probability is not resolved, then the above Lemma 2.9 can be simplified in a way that the resulting quantitative information no longer depends on associated moduli of absolute continuity. Concretely, suppose that (X_n) satisfies

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n + \frac{\varepsilon}{N}$$

for all $n < N$, that is (X_n) is a λ - (ε/N) - N -supermartingale w.r.t. (\mathcal{F}_n) for any $\lambda > 0$. Then

$$\mathbb{E}[X_\tau \mid \mathcal{F}_\rho] \leq X_\rho + \varepsilon.$$

To see this, write $X_n = Y_n + Z_n$ with (Y_n) and (Z_n) as in Doob's decomposition theorem. Then, as (X_n) is a λ - (ε/N) - N -supermartingale for any λ , we get that (Z_n) is λ - (ε/N) - N -nonincreasing for any $\lambda > 0$, by Lemma 2.3. By Doob's optional stopping theorem for martingales, we get

$$\mathbb{E}[Y_\tau \mid \mathcal{F}_\rho] = Y_\rho$$

as (Y_n) is a martingale. Thus, we get

$$\mathbb{E}[X_\tau \mid \mathcal{F}_\rho] = Y_\rho + \mathbb{E}[Z_\tau \mid \mathcal{F}_\rho]$$

As $Z_{n+1} \leq Z_n + \varepsilon/N$ for any $n < N$, we have $Z_\rho + \varepsilon \geq Z_\tau$ and so in particular

$$\mathbb{E}[Z_\rho \mid \mathcal{F}_\rho] + \varepsilon \geq \mathbb{E}[Z_\tau \mid \mathcal{F}_\rho].$$

As $Z_\rho = \mathbb{E}[Z_\rho \mid \mathcal{F}_\rho]$, one derives

$$\mathbb{E}[X_\tau \mid \mathcal{F}_\rho] \leq X_\rho + \varepsilon$$

which was the claim.

Note that this type of strengthening, under stronger approximate martingale properties, can also be observed for all preceding results, and for various of the following results. However, contrary to the preceding results, the optional stopping theorem is the first result where this special case presented above is not a corollary of our approach, which is why we explicitly remark it here.

For two processes $C := (C_n)$ and $X := (X_n)$, we define the discrete stochastic integral $C \bullet X$ as usual via

$$(C \bullet X)_n := \sum_{i=1}^n C_i(X_i - X_{i-1}).$$

The following is a variant for finitary supermartingales of the fundamental result about the discrete stochastic integral (c.f. [31, Theorem 9.39]) that if C is a nonnegative and bounded predictable process and X is a supermartingale, then $C \bullet X$ is also a supermartingale.

Lemma 2.11. *Let $C := (C_n)$ be a bounded, nonnegative and predictable process and let $K > 0$ be such that $C_n \leq K$ for any $n \in \mathbb{N}$. Further, let $X := (X_n)$ be a λ - (ε/K) - N -supermartingale. Then $C \bullet X$ is a λ - ε - N -supermartingale.*

Proof. Write $Y_n := (C \bullet X)_n$. Then we have by definition that

$$\mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{F}_n] = C_{n+1}(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n).$$

Let now ω be such that $C_n(\omega) \leq K$ as well as $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n](\omega) \leq X_n(\omega) + \varepsilon/K$. Then we have

$$\mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{F}_n](\omega) = C_{n+1}(\omega)(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n](\omega) - X_n(\omega)) \leq C_{n+1}(\omega)\varepsilon/K \leq \varepsilon.$$

Then we have

$$\mathbb{P}(\mathbb{E}[Y_{n+1} - Y_n \mid \mathcal{F}_n] > \varepsilon) \leq \mathbb{P}(C_{n+1} > K) + \mathbb{P}(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] - X_n > \varepsilon/K) < \lambda$$

which completes the proof. \square

In another result, we have that we can convert moduli of absolute continuity from elements of a nonnegative stochastic process to a discrete stochastic integral.

Lemma 2.12. *Let $C := (C_n)$ be a bounded, nonnegative and predictable process and let $K > 0$ be such that $C_n \leq K$ for any $n \in \mathbb{N}$. Further, let $X := (X_n)$ be a nonnegative stochastic process with moduli of absolute continuity (μ_n) . Then $\mu_N^M(\varepsilon/KN) := \min\{\mu_n(\varepsilon/KN) \mid n \leq N\}$ is a modulus of absolute continuity for $(C \bullet X)_n$ for all $n \leq N$.*

Proof. Given $n \leq N$, we have

$$\begin{aligned} \mathbb{E}[(C \bullet X)_n \mathbf{1}_A] &= \sum_{i=1}^n \mathbb{E}[C_i(X_i - X_{i-1}) \mathbf{1}_A] \\ &\leq K \sum_{i=1}^n \mathbb{E}[X_i \mathbf{1}_A] \\ &\leq K \sum_{i=1}^n \varepsilon / KN \leq \varepsilon \end{aligned}$$

whenever A is such that $\mathbb{P}(A) < \mu_N^M(\varepsilon / KN) := \min\{\mu_n(\varepsilon / KN) \mid n \leq N\}$. \square

The final result we need before stating our quantitative martingale convergence theorem is a downcrossing inequality for finitary supermartingales. Crossing inequalities, typically involving *upcrossings*, are the standard way of establishing martingale convergence. For our quantitative version, downcrossings offer a slightly more direct route in the case of supermartingales, as previously illustrated in the case of almost-supermartingales in [45, Remark 7.11], which is in turn a variant of standard downcrossing inequalities as presented in e.g. [21]. We first recall the notion of a downcrossing:

Definition 2.13. Given $N \in \mathbb{N}$, $\alpha < \beta$ and a sequence of random variables (X_n) , we denote by $D_{N, [\alpha, \beta]}(X_n)$ the number of times the process (X_n) downcrosses the interval $[\alpha, \beta]$ before time N . More precisely, $D_{N, [\alpha, \beta]}(X_n)(\omega) = k$ for the maximal $k \in \mathbb{N}$ such that there exists a sequence $i_1 < j_1 \leq i_2 < j_2 \leq \dots \leq i_k < j_k < N$ such that $\beta \leq X_{i_l}(\omega)$ and $X_{j_l}(\omega) \leq \alpha$ for all $l = 1, \dots, k$.

Our quantitative downcrossing inequality, phrased for finitary supermartingales, then takes the following form:

Lemma 2.14. *Let $\varepsilon > 0$, $N \in \mathbb{N}$ and $\alpha < \beta$ be given with $0 < \delta \leq \beta - \alpha$. Let (X_n) be a nonnegative stochastic process with moduli of absolute continuity (μ_n) . If (X_n) is a $\mu_N^M(\varepsilon\delta/4N^2)$ - $(\varepsilon\delta/2N)$ - N -supermartingale, then*

$$\mathbb{E}[D_{N, [\alpha, \beta]}(X_n)] \leq \frac{1}{\beta - \alpha} (\mathbb{E}[X_0 - X_N] + \mathbb{E}[(X_N - \beta)^+]) + \varepsilon.$$

Here, as before: $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$. In case that $\mathbb{E}[X_0] \leq K$, the inequality simplifies to

$$\mathbb{E}[D_{N, [\alpha, \beta]}(X_n)] \leq \frac{1}{\beta - \alpha} (K + \mathbb{E}[(X_N - \beta)^+]) + \varepsilon.$$

Proof. Given $\alpha < \beta$, define the $\{0, 1\}$ -valued predictable process $C := (C_n)$ by $C_0 := 1$ iff $X_0 \geq \beta$ and

$$C_{n+1} := 1 \text{ iff } \begin{cases} C_n = 1 \text{ and } X_n > \alpha, \text{ or} \\ C_n = 0 \text{ and } X_n \geq \beta. \end{cases}$$

Let $C' := (C'_n)$ be the dual predictable process defined by $C'_n := 1 - C_n$. We have

$$(\dagger) \quad (\beta - \alpha) D_{N, [\alpha, \beta]}(X_n) \leq - \sum_{i=1}^N C_i (X_i - X_{i-1}) + (X_N - \beta)^+$$

pointwise everywhere, the proof of which we defer to the end. For now assuming (\dagger) , since C' is a nonnegative predictable process, bounded by 1 and (X_n) is a $\mu_N^M(\varepsilon\delta/4N^2)$ - $(\varepsilon\delta/2N)$ - N -supermartingale, we must have that $C' \bullet X$ is a $\mu_N^M(\varepsilon\delta/4N^2)$ - $(\varepsilon\delta/2N)$ - N -supermartingale by

Lemma 2.11, where further $(C' \bullet X)_0 = 0$. Lemma 2.12 yields that $\mu_N^M(\varepsilon/N)$ is a modulus of absolute continuity for $(C' \bullet X)_n$ for all $n \leq N$. Therefore, Lemma 2.6 yields

$$\mathbb{E} \left[\sum_{i=1}^N C'_i(X_i - X_{i-1}) \right] = \mathbb{E}[(C' \bullet X)_N] \leq \mathbb{E}[(C' \bullet X)_0] + \varepsilon\delta \leq \varepsilon(\beta - \alpha).$$

This in turn yields

$$\begin{aligned} \mathbb{E} \left[-\sum_{i=1}^N C_i(X_i - X_{i-1}) \right] - \varepsilon(\beta - \alpha) &\leq \mathbb{E} \left[-\sum_{i=1}^N C_i(X_i - X_{i-1}) \right] + \mathbb{E} \left[-\sum_{i=1}^N C'_i(X_i - X_{i-1}) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^N (C_i + C'_i)(X_{i-1} - X_i) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^N (X_{i-1} - X_i) \right] \\ &= \mathbb{E}[X_0 - X_N] \end{aligned}$$

and so it follows from (\dagger) that

$$(\beta - \alpha)\mathbb{E}[D_{N, [\alpha, \beta]}(X_N)] \leq \mathbb{E}[X_0 - X_N] + \mathbb{E}[(X_N - \beta)^+] + \varepsilon(\beta - \alpha)$$

and the result follows immediately.

It remains to show (\dagger) . To see this, fix $\omega \in \Omega$, set $k := D_{N, [\alpha, \beta]}(X_n(\omega))$, and take a sequence

$$i_1 < j_1 \leq i_2 < j_2 \leq \dots \leq i_k < j_k < N \text{ with } \beta \leq X_{i_l}(\omega) \text{ and } X_{j_l}(\omega) \leq \alpha,$$

where furthermore we assume (w.l.o.g.) that $i_l \geq j_{l-1}$ is the first such index with $\beta \leq X_{i_l}(\omega)$, and similarly $j_l > i_l$ the first with $X_{j_l}(\omega) \leq \alpha$. It is easy to see by induction that for $1 \leq p \leq j_k$, $C_p(\omega) = 1$ iff $i_l < p \leq j_l$ for some $1 \leq l \leq k$. So, in particular $C_{j_k}(\omega) = 1$ and so $C_{j_k+1}(\omega) = 0$. Now, if $X_p(\omega) < \beta$ for all $N \geq p > j_k$, then for all such p we must have $C_p(\omega) = 0$ and so

$$\begin{aligned} -\sum_{i=1}^N C_i(\omega)(X_i(\omega) - X_{i-1}(\omega)) &= -\sum_{l=1}^k \sum_{p=i_l+1}^{j_l} (X_p(\omega) - X_{p-1}(\omega)) \\ &= \sum_{l=1}^k X_{i_l}(\omega) - X_{j_l}(\omega) \geq (\beta - \alpha)D_{N, [\alpha, \beta]}(X_n(\omega)) \end{aligned}$$

and we obtain the desired inequality, noting that $(X_N(\omega) - \beta)^+ = 0$ in this case. On the other hand, if there is some $N \geq p > j_k$ satisfying $X_p(\omega) \geq \beta$, take the first such p . This would imply that $C_q(\omega) = 1$ for all $N \geq q > p$ (by maximality of k) and $C_q(\omega) = 0$ for $j_k < q \leq p$. This yields

$$\begin{aligned} -\sum_{i=1}^N C_i(\omega)(X_i(\omega) - X_{i-1}(\omega)) &= -\sum_{l=1}^k \sum_{r=i_l+1}^{j_l} (X_r(\omega) - X_{r-1}(\omega)) - \sum_{q=p+1}^N (X_q(\omega) - X_{q-1}(\omega)) \\ &= \sum_{l=1}^k (X_{i_l}(\omega) - X_{j_l}(\omega)) + X_p(\omega) - X_N(\omega). \end{aligned}$$

The above and the fact that $X_p(\omega) \geq \beta$ yields

$$-\sum_{i=1}^N C_i(\omega)(X_i(\omega) - X_{i-1}(\omega)) + X_N(\omega) - \beta \geq (\beta - \alpha)D_{N, [\alpha, \beta]}(X_n(\omega))$$

and the desired inequality holds on Ω . \square

We now state and prove the main result of the section: a quantitative version of Doob's L_1 -martingale convergence theorem for the special case of nonnegative supermartingales. While it is an immediate consequence of the martingale convergence theorem that any nonnegative supermartingale converges almost surely to some X with $\mathbb{E}[|X|] < +\infty$, in general, producing a computable rate of convergence is generally not possible (as discussed in the introduction). As such, and following [45], we instead provide a rate of metastability for the corresponding Cauchy property, in the simplified form of a learnable rate of uniform convergence (cf. [45, Theorem 7.4]). The main difference here is that our quantitative result is now given in terms of finitary supermartingales. For further background on concepts such as uniform learnability and their relationship to martingale convergence, the reader is directed to [45]. For the purpose of this paper, the importance of this result is that it is used in the next section to establish fluctuations bounds for nonnegative supermartingales (Theorem 3.5), which are in turn used to obtain fluctuation bounds for the Robbins-Siegmund theorem (Theorem 3.9), which represents a crucial component of our overall analysis of the convergence of stochastic quasi-Fejér monotone sequences.

Theorem 2.15. *For any $\varepsilon, \lambda \in (0, 1]$ and $K > 0$, define*

$$N_K(\lambda, \varepsilon) := \left\lceil \frac{512(K+1)^2}{\lambda^2 \varepsilon^2} \right\rceil.$$

Let (X_n) be a nonnegative stochastic process with moduli of absolute continuity (μ_n) . Then, if (X_n) is a $\mu_N^M(\varepsilon^2/32N_K^2(\lambda, \varepsilon))$ - $(\varepsilon^2/16N_K(\lambda, \varepsilon))$ - $N_K(\lambda, \varepsilon)$ -supermartingale with $\mathbb{E}[X_0] < K$, there exists an $i \leq N_K(\lambda, \varepsilon)$ such that

$$\mathbb{P}(\exists k, l \in [a_i; b_i] (|X_k - X_l| \geq \varepsilon)) < \lambda,$$

uniformly over any $a_0 < b_0 \leq a_1 < b_1 \leq \dots < N_K(\lambda, \varepsilon)$. Here, as before: $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$.

Proof. Fix ε, λ and K and let (X_n) be a nonnegative $\mu_N^M(\varepsilon^2/32N_K^2(\lambda, \varepsilon))$ - $(\varepsilon^2/16N_K(\lambda, \varepsilon))$ - $N_K(\lambda, \varepsilon)$ -supermartingale with $\mathbb{E}[X_0] < K$. First observe that, by (the proof of) Lemma 2.6, we have

$$\mathbb{E}[X_n] \leq \mathbb{E}[X_0] + \varepsilon/8 < K + 1$$

for each $n \leq N_K(\lambda, \varepsilon)$, as $\varepsilon \leq 1$. Now define the events

$$A_i := \{\omega \mid \exists k, l \in [a_i; b_i] (|X_k(\omega) - X_l(\omega)| \geq \varepsilon)\} \quad \text{and} \quad B_i := \{\omega \mid X_{a_i}(\omega) \leq 2(K+1)/\lambda\}$$

for $i = 0, \dots, N_K(\lambda, \varepsilon)$. Divide the interval $[0, 2(K+1)/\lambda]$ into $p := \lceil 8(K+1)/\lambda\varepsilon \rceil$ many equal subintervals which we label $[\alpha_j, \beta_j]$ for $j = 1, \dots, p$. Add another subinterval of the same width on either side of $[0, 2(K+1)/\lambda]$, which we label $[\alpha_0, \beta_0]$ and $[\alpha_{p+1}, \beta_{p+1}]$ respectively. We note that each subinterval has width $\leq \varepsilon/4$.

Now, if $\omega \in A_i \cap B_i$ there exist $k(\omega), l(\omega) \in [a_i; b_i]$ with $|X_{k(\omega)}(\omega) - X_{l(\omega)}(\omega)| \geq \varepsilon$. Then by the triangle inequality, either $|X_{a_i}(\omega) - X_{k(\omega)}(\omega)| \geq \varepsilon/2$ or $|X_{a_i}(\omega) - X_{l(\omega)}(\omega)| \geq \varepsilon/2$, and since also $X_{a_i}(\omega) \leq 2(K+1)/\lambda$, this means that one of the intervals $[\alpha_j, \beta_j]$ for $j = 0, \dots, p+1$ is upcrossed or downcrossed somewhere in $[a_i; b_i]$. Therefore, defining the event

$$C_{i,j} := \text{"}(X_n) \text{ crosses } [\alpha_j, \beta_j] \text{ somewhere in } [a_i; b_i]\text{"},$$

we have shown that

$$A_i \cap B_i \subseteq \bigcup_{j=0}^{p+1} C_{i,j}$$

and this yields

$$\sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbb{P}(A_i \cap B_i) \leq \sum_{i=0}^{N_K(\lambda,\varepsilon)} \sum_{j=0}^{p+1} \mathbb{P}(C_{i,j}).$$

For any $j \in \{0, \dots, p+1\}$, we further have

$$\begin{aligned} \sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbb{P}(C_{i,j}) &= \sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbb{E}[\mathbf{1}_{C_{i,j}}] = \mathbb{E} \left[\sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbf{1}_{C_{i,j}} \right] \\ &\leq 2\mathbb{E}[D_{N_K(\lambda,\varepsilon),[\alpha_j,\beta_j]}(X_n)] + 1, \end{aligned}$$

where for the final inequality we note that the quantity $\sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbf{1}_{C_{i,j}}$ is bounded by the total number of times that (X_n) crosses $[\alpha_j, \beta_j]$ within its initial segment of length $N_K(\lambda, \varepsilon)$, and since between every two downcrossings there exists exactly one upcrossing, the total number of crossings (being the sum of the number of downcrossings and upcrossings) is bounded by $2D_{N_K(\lambda,\varepsilon),[\alpha_j,\beta_j]}(X_n) + 1$. Applying Lemma 2.14, we see that

$$\mathbb{E}[D_{N_K(\lambda,\varepsilon),[\alpha_j,\beta_j]}(X_n)] \leq \frac{1}{\beta_j - \alpha_j} (K + \mathbb{E}[(X_{N_K(\lambda,\varepsilon)} - \beta_j)^+]) + 1$$

since (X_n) is an $\mu_N^M(\varepsilon^2/32N_K^2(\lambda, \varepsilon)) - (\varepsilon^2/16N_K(\lambda, \varepsilon)) - N_K(\lambda, \varepsilon)$ -supermartingale by assumption and we can show that $\beta_j - \alpha_j \geq \varepsilon/8$ for any j . We have

$$\frac{1}{\beta_j - \alpha_j} (K + \mathbb{E}[(X_{N_K(\lambda,\varepsilon)} - \beta_j)^+]) + 1 \leq \frac{8(K + \mathbb{E}[X_{N_K(\lambda,\varepsilon)}])}{\varepsilon} + 1 \leq \frac{8(2K + 1) + 1}{\varepsilon}$$

using that $\beta_j \geq 0$ for all j (and $\varepsilon \leq 1$). Picking a $j_0 \in \{0, \dots, p+1\}$ where $\mathbb{P}(C_{i,j_0})$ is maximal among all $\mathbb{P}(C_{i,j})$, we now get

$$\sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbb{P}(A_i \cap B_i) \leq (p+2) \sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbb{P}(C_{i,j_0})$$

for this j_0 and so, combined with the above and noting $\varepsilon, \lambda \in (0, 1]$, we get

$$\begin{aligned} \sum_{i=0}^{N_K(\lambda,\varepsilon)} \mathbb{P}(A_i \cap B_i) &\leq (p+2)(2\mathbb{E}[D_{N_K(\lambda,\varepsilon),[\alpha_{j_0},\beta_{j_0}]}(X_n)] + 1) \\ &< \left(\frac{8(K+1)}{\lambda\varepsilon} + 3 \right) \left(\frac{16(2K+1)+2}{\varepsilon} + 1 \right) \\ &\leq \frac{256(K+1)^2}{\lambda\varepsilon^2}. \end{aligned}$$

Therefore there is some $i \in \{0, \dots, N_K(\lambda, \varepsilon)\}$ such that

$$\mathbb{P}(A_i \cap B_i) < \frac{1}{N_K(\lambda, \varepsilon) + 1} \frac{256(K+1)^2}{\lambda\varepsilon^2} \leq \frac{\lambda^2\varepsilon^2}{512(K+1)^2} \frac{256(K+1)^2}{\lambda\varepsilon^2} = \frac{\lambda}{2}.$$

But then, using Boole's inequality, it follows that

$$\mathbb{P}(A_i) + \mathbb{P}(B_i) - 1 \leq \mathbb{P}(A_i \cap B_i) < \frac{\lambda}{2}$$

and by Markov's inequality and the fact that $a_i < N_K(\lambda, \varepsilon)$, we have

$$\mathbb{P}(B_i) = 1 - \mathbb{P}(B_i^c) = 1 - \mathbb{P}\left(X_{a_i} > \frac{2(K+1)}{\lambda}\right) \geq 1 - \frac{\mathbb{E}[X_{a_i}]}{2(K+1)/\lambda} > 1 - \frac{\lambda}{2}$$

so that $\mathbb{P}(A_i) < \lambda$. □

3. A QUANTITATIVE ROBBINS-SIEGMUND THEOREM

We now develop our theory of finitary martingales to give a quantitative result on almost-supermartingale convergence, more specifically a quantitative version of the classic Robbins-Siegmund theorem [52], which forms a central part of our main analysis just as the (qualitative) Robbins-Siegmund theorem forms a crucial part of the study of stochastic quasi-Fejér monotone sequences (see in particular [17]). Reflecting the fundamentally pointwise nature of the convergence proofs for such sequences in (locally) compact spaces, such as in [17], the main result of this paper is a so-called metastable rate of pointwise convergence for stochastic quasi-Fejér monotone sequences, as previously discussed in the introduction. As shown in [45], metastable rates of pointwise convergence take on a particularly simple form when we are in possession of suitable quantitative information on the fluctuation behaviour of the sequence, and though the complexity of our overall argument, notably the use of compactness, precludes our final rate having such a simple form so that we are only able to derive such a metastable rate, we focus on fluctuation behaviour to the greatest extent possible in order to simplify our analysis. For the Robbins-Siegmund theorem in particular, quantitative information on fluctuations in the ordinary sense is possible, and it is the goal of this section to characterise this information. While we broadly follow the approach of the recent quantitative analysis of the Robbins-Siegmund theorem in [46], the main fluctuation bound presented in Theorem 3.9 crucially dispenses of the so-called uniformly learnable rates derived therein and replaces them by fluctuation bounds. In that way, the bounds presented in Theorem 3.9 should actually be viewed as an extension to L_1 -bounded λ - ε - N -almost-supermartingales of the much older work of Chashka [13] and Kachurovskii [30] on fluctuations in martingales.

To begin, for the benefit of the reader, we recall the notion of fluctuations here:

Definition 3.1. Given $\varepsilon > 0$, $N \in \mathbb{N}$ and a sequence of random variables (X_n) , we denote by $J_{N,\varepsilon}(X_n)$ the number of ε -fluctuations experienced by (X_n) before time N . More precisely, $J_{N,\varepsilon}(X_n)(\omega) = k$ for the maximal $k \in \mathbb{N}$ such that there exists a sequence $i_1 < j_1 \leq i_2 < j_2 \leq \dots \leq i_k < j_k < N$ such that $|X_{i_l}(\omega) - X_{j_l}(\omega)| \geq \varepsilon$ for all $l = 1, \dots, k$.

We first collect the necessary minor results we need for our main task. The following lemma outlines some standard inequalities relating fluctuations and various (arithmetical) operations on stochastic processes:

Lemma 3.2. *Let (X_n) , (Y_n) be sequences of random variables. Then for any $a, b, c > 0$, any $\varepsilon > 0$ and any $n \in \mathbb{N}$:*

- (i) $\mathbb{P}(\max_{i \leq n} |X_i + Y_i| \geq a) \leq \mathbb{P}(\max_{i \leq n} |X_i| \geq a/2) + \mathbb{P}(\max_{i \leq n} |Y_i| \geq a/2)$,
- (ii) $\mathbb{P}(\max_{i \leq n} |X_i Y_i| \geq a) \leq \mathbb{P}(\max_{i \leq n} |X_i| \geq \sqrt{a}) + \mathbb{P}(\max_{i \leq n} |Y_i| \geq \sqrt{a})$,
- (iii) $\mathbb{P}(J_{n,\varepsilon}(X_i + Y_i) \geq a) \leq \mathbb{P}(J_{n,\varepsilon/2}(X_i) \geq a/2) + \mathbb{P}(J_{n,\varepsilon/2}(Y_i) \geq a/2)$,
- (iv) $\mathbb{P}(J_{n,\varepsilon}(X_i Y_i) \geq a) \leq \mathbb{P}(J_{n,\varepsilon/c}(X_i) \geq a/2) + \mathbb{P}(J_{n,\varepsilon/b}(Y_i) \geq a/2) + \mathbb{P}(\max_{i \leq n} |X_i| \geq b) + \mathbb{P}(\max_{i \leq n} |Y_i| \geq c)$.

Proof. All bounds are rather immediate. The most involved is (iv), for which we provide a proof as an example: Write $x_i := X_i(\omega)$ and $y_i := Y_i(\omega)$ and assume that $\max_{i \leq n} |x_i| < b$ and $\max_{i \leq n} |y_i| < c$. If we have a fluctuation $|x_i y_i - x_j y_j| \geq \varepsilon$, then

$$|y_i||x_i - x_j| + |x_j||y_i - y_j| \geq \varepsilon$$

and thus either

$$|x_i - x_j| \geq \varepsilon/|y_i| > \varepsilon/c \quad \text{or} \quad |y_i - y_j| \geq \varepsilon/|x_j| > \varepsilon/b.$$

In other words, any single ε -fluctuation of the sequence $(x_k y_k)$ corresponds to an ε/c fluctuation of (x_k) or an ε/b -fluctuation of (y_k) , and thus

$$J_{n,\varepsilon}(x_k y_k) \leq J_{n,\varepsilon/c}(x_k) + J_{n,\varepsilon/b}(y_k).$$

We have therefore shown that

$$\left\{ J_{n,\varepsilon}(X_i Y_i) \geq a \cap \sup_{i \leq n} |X_i| < b \cap \sup_{i \leq n} |Y_i| < c \right\} \subseteq \left\{ J_{n,\varepsilon/c}(X_i) \geq a/2 \cup J_{n,\varepsilon/b}(Y_i) \geq a/2 \right\}$$

and the inequality follows. \square

Beyond fluctuations, we will also require some further quantitative results on finitary martingales in the following, starting with a quantitative version of the standard fact that stopped supermartingales are also supermartingales (cf. [31, Theorem 10.15]).

Lemma 3.3. *Let τ be a stopping time w.r.t. (\mathcal{F}_n) . If (X_n) is a λ - ε - N -supermartingale w.r.t. (\mathcal{F}_n) , so is the stopped process $(X_{n \wedge \tau})$, where $n \wedge \tau := \min\{n, \tau\}$.*

Proof. Fixing $n < N$, by basic properties of conditional expectations and stopping times, we have

$$\begin{aligned} \mathbb{E}[X_{(n+1) \wedge \tau} \mid \mathcal{F}_n] &= \mathbb{E} \left[\sum_{i=0}^n X_i \mathbf{1}_{\{\tau=i\}} \mid \mathcal{F}_n \right] + \mathbb{E} [X_{n+1} \mathbf{1}_{\{\tau \geq n+1\}} \mid \mathcal{F}_n] \\ &= \sum_{i=0}^n X_i \mathbf{1}_{\{\tau=i\}} + \mathbf{1}_{\{\tau \geq n+1\}} \mathbb{E} [X_{n+1} \mid \mathcal{F}_n], \end{aligned}$$

on some set Ω' of measure one. Now, let $\omega \in \Omega'$ be such that $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n](\omega) \leq X_n(\omega) + \varepsilon$. Then the above yields

$$\begin{aligned} \mathbb{E}[X_{(n+1) \wedge \tau} \mid \mathcal{F}_n](\omega) &\leq \sum_{i=0}^n X_i(\omega) \mathbf{1}_{\{\tau=i\}}(\omega) + \mathbf{1}_{\{\tau \geq n+1\}}(\omega) X_n(\omega) + \mathbf{1}_{\{\tau \geq n+1\}}(\omega) \varepsilon \\ &= \sum_{i=0}^{n-1} X_i(\omega) \mathbf{1}_{\{\tau=i\}}(\omega) + \mathbf{1}_{\{\tau \geq n\}}(\omega) X_n(\omega) + \mathbf{1}_{\{\tau \geq n+1\}}(\omega) \varepsilon \\ &\leq X_{n \wedge \tau}(\omega) + \varepsilon. \end{aligned}$$

In particular, we have shown that

$$\mathbb{P}(\mathbb{E}[X_{(n+1) \wedge \tau} \mid \mathcal{F}_n] > X_{n \wedge \tau} + \varepsilon) \leq \mathbb{P}(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] > X_n + \varepsilon)$$

which yields the claim. \square

The next results presents our finitary variant of a classic inequality of Ville [56] (see also [40]).

Lemma 3.4. *Let (X_n) be a nonnegative stochastic process with moduli of absolute continuity (μ_n) . If (X_n) is a $\mu_N^M(\varepsilon/8N^2)$ - $\varepsilon/2N$ - N -supermartingale, then for any $\alpha > 0$ and any $k \leq N$, we have*

$$\mathbb{P}(\exists t \in [0; k] (X_t \geq \alpha)) \leq \frac{\mathbb{E}[X_0]}{\alpha} + \frac{\varepsilon}{\alpha}.$$

Here, as before: $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$.

Proof. We define the stopping time $\tau = \inf\{t \leq N \mid X_t \geq \alpha\}$. By Lemma 2.8, we have that $\mu_N^M(\varepsilon/N)$ is a modulus of absolute continuity for X_τ . In particular, as $X_{k \wedge \tau} \leq X_k + X_\tau$, we therefore have that $\mu_N^M(\varepsilon/2N)$ is a modulus of absolute continuity for $X_{k \wedge \tau}$. By Lemma 3.3, $X_{k \wedge \tau}$ is a $\mu_N^M(\varepsilon/8N^2)$ - $\varepsilon/2N$ - N -supermartingale. Lemma 2.6 thereby yields that

$$\mathbb{E}[X_{k \wedge \tau}] \leq \mathbb{E}[X_0] + \varepsilon$$

as $k \wedge \tau \leq k$. Therefore, since if $\tau(\omega) \leq k$, then $X_{k \wedge \tau(\omega)}(\omega) = X_{\tau(\omega)}(\omega) \geq \alpha$, it follows that

$$\begin{aligned} \mathbb{P}(\exists t \in [0; k] (X_t \geq \alpha)) &= \mathbb{P}(\tau \leq k) \\ &\leq \mathbb{P}(X_{k \wedge \tau} \geq \alpha) \\ &\leq \frac{\mathbb{E}[X_{k \wedge \tau}]}{\alpha} \\ &\leq \frac{\mathbb{E}[X_0]}{\alpha} + \frac{\varepsilon}{\alpha}, \end{aligned}$$

where we used Markov's inequality for $X_{k \wedge \tau}$. \square

The first main result of this section is to provide a so-called modulus of finite fluctuations (following the terminology of [45, Definition 4.5]) for nonnegative supermartingales, now adapted to the finitary setting. It can be shown (cf. Section 4.2 of [45]) that if (X_n) converges, then for any $\varepsilon > 0$ we have

$$\lim_{k \rightarrow \infty} \mathbb{P} \left(\lim_{N \rightarrow \infty} J_{N, \varepsilon}(X_n) \geq k \right) = 0.$$

A modulus of finite fluctuations in the ordinary sense is defined to be (essentially) a rate of convergence for the above property parametrised by $\varepsilon > 0$, namely a function $\phi : (0, \infty)^2 \rightarrow \mathbb{R}$ satisfying

$$\mathbb{P}(J_{N, \varepsilon}(X_n) \geq \phi(\lambda, \varepsilon)) < \lambda$$

for all $\lambda, \varepsilon > 0$ and $N \in \mathbb{N}$. Such convergence rates provide interesting quantitative information on the oscillatory characteristics of processes, and have been widely studied in the context of both probability theory and ergodic theory (see in particular [30]). Indeed, explicit moduli of finite fluctuations for L_p -bounded martingales are already provided in [13] and later in [29]. In the following result we generalise these to the case of finitary supermartingales, combining elements of the proof of [13, Theorem 1] with our quantitative result on finitary supermartingale convergence (Theorem 2.15). We note that an alternative approach here would have been to instead attempt to adapt the entire set of proofs in [13] or [29] to manufacture finitary versions, though it is not obvious how these can be extended to supermartingales rather than just martingales, and in any case we conjecture that the improvement in the approximation bound, if any, would be minor.

Theorem 3.5. *Fix $\lambda, \varepsilon > 0$ and $N \in \mathbb{N}$. Let (X_n) be a nonnegative stochastic process with moduli of absolute continuity (μ_n) . Let (X_n) be a $\mu_N^M(\varepsilon_0 \lambda_0 / 16N)$ - $(\varepsilon_0 \lambda_0 / 8N)$ - N -supermartingale w.r.t. (\mathcal{F}_n) , where $\lambda_0 := \min\{\mu_e^M(\varepsilon^2 / 128e^2) \mid e \leq N\}$ and $\varepsilon_0 := \varepsilon^2 / 64N$. Let $K > 0$ be such that $\mathbb{E}[X_0] < K$. Then, for any $\lambda > 0$, we have*

$$\mathbb{P} \left(J_{N, \varepsilon}(X_n) \geq \left\lceil \frac{2048(K+1)^2}{\lambda^2 \varepsilon^2} \right\rceil \right) < \lambda.$$

Here, as before, we use the notation $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$.

Proof. We define the stopping times (τ_j) as follows: We set $\tau_0 := 0$ and define

$$\tau_j := \begin{cases} \inf\{i \in [\tau_{j-1}; N] \mid |X_{\tau_{j-1}} - X_i| \geq \varepsilon/2\} & \text{if existent,} \\ N & \text{otherwise.} \end{cases}$$

Now, setting

$$e := \left\lceil \frac{2048(K+1)^2}{\lambda^2 \varepsilon^2} \right\rceil,$$

and using that fact that if, for any point $m \in \mathbb{N}$, (X_n) experiences an ε -fluctuation after time m , then there is some point $l \geq m$ such that $|X_m - X_l| \geq \varepsilon/2$, we see that

$$\mathbb{P}(J_{N,\varepsilon}(X_n) \geq e) \leq \mathbb{P}\left(\bigcap_{i \leq e} (|X_{\tau_i} - X_{\tau_{i+1}}| \geq \varepsilon/2)\right).$$

Since $\tau_j \leq N$ for all $j \in \mathbb{N}$, by Lemma 2.9, we have that (X_{τ_n}) is a nonnegative λ_0 - ε_0 - N -supermartingale with respect to the filtration (\mathcal{F}_{τ_n}) with $\mathbb{E}[X_{\tau_0}] = \mathbb{E}[X_0] < K$. We now apply Theorem 2.15, noting that we can assume $e \leq N$ (otherwise the theorem is trivially true) and therefore that (X_{τ_n}) being a λ_0 - ε_0 - N -supermartingale implies that it is also an $\mu_e^M(\varepsilon^2/128e^2)$ - $(\varepsilon^2/64e)$ - e -supermartingale. Taking $a_i := i$ and $b_i := i + 1$, there thereby exists an $i_0 \leq e$ satisfying

$$\mathbb{P}(|X_{\tau_{i_0}} - X_{\tau_{i_0+1}}| \geq \varepsilon/2) = \mathbb{P}(\exists k, l \in [i_0; i_0 + 1] \mid |X_{\tau_k} - X_{\tau_l}| \geq \varepsilon/2) < \lambda.$$

The result follows by noting that then

$$\mathbb{P}(J_{N,\varepsilon}(X_n) \geq e) \leq \mathbb{P}\left(\bigcap_{i \leq e} (|X_{\tau_i} - X_{\tau_{i+1}}| \geq \varepsilon/2)\right) \leq \mathbb{P}(|X_{\tau_{i_0}} - X_{\tau_{i_0+1}}| \geq \varepsilon/2) < \lambda. \quad \square$$

Remark 3.6. Akin to Remark 2.10, if (X_n) is already a finitary martingale in the stronger sense that the associated probability is not resolved, then the above Theorem 3.5 can be simplified in a way that the resulting quantitative information no longer depends on associated moduli of absolute continuity. The proof then in particular relies on the more uniform variant of the quantitative optional stopping theorem established in that remark.

Our last preliminary quantitative result before we move to the Robbins-Siegmund theorem is the following:

Lemma 3.7. *If (X_n) is a monotone, nonnegative sequence of random variables with*

$$\mathbb{P}\left(\sup_{n \in \mathbb{N}} X_n \geq f(\lambda)\right) < \lambda$$

for all $\lambda > 0$ and some given function $f : (0, \infty) \rightarrow \mathbb{N}$, then we have

$$\mathbb{P}\left(J_{n,\varepsilon}(X_i) \geq \frac{2f(\lambda)}{\varepsilon}\right) < \lambda$$

for all $n \in \mathbb{N}$ and $\varepsilon, \lambda > 0$.

Proof. Fix $n \in \mathbb{N}$ as well as $\varepsilon, \lambda > 0$. First we have for any $a > 0$ that

$$\mathbb{P}(J_{n,\varepsilon}(X_i) \geq a) < \mathbb{P}\left(J_{n,\varepsilon}(X_i) \geq a \cap \max_{i \leq n} X_i < f(\lambda)\right) + \lambda$$

Now, let $J_{n,\varepsilon}(X_i(\omega)) \geq a$ and $\max_{i \leq n} X_i(\omega) < f(\lambda)$ for $a := 2f(\lambda)/\varepsilon > \lceil f(\lambda)/\varepsilon \rceil$. Since the sequence $(X_i(\omega))_{i \leq n}$ is monotone and lies in $[0, f(\lambda)]$, there can be at most $\lceil f(\lambda)/\varepsilon \rceil$ -many ε -fluctuations and so we have

$$\mathbb{P} \left(J_{n,\varepsilon}(X_i) \geq a \cap \max_{i \leq n} X_i < f(\lambda) \right) = 0$$

which yields the result. \square

We now come to the main result of the section, and the culmination of our development of finitary martingale theory. The well-known Robbins-Siegmund theorem [52] is an extension of Doob's convergence theorem to certain almost supermartingales, stating that whenever (X_n) is a nonnegative process adapted to (\mathcal{F}_n) such that

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq (1 + \chi_n)X_n + \eta_n - \zeta_n$$

for parameter sequences (χ_n) , (η_n) , (ζ_n) with $\sum_{i=0}^{\infty} \chi_i < +\infty$ and $\sum_{i=0}^{\infty} \eta_i < +\infty$, then it holds that (X_n) converges to a finite limit and $\sum_{i=0}^{\infty} \zeta_i < +\infty$. In what follows we do not require the second part of the Robbins-Siegmund theorem and we thus restrict our attention to the special case that $\zeta_n = 0$. However, akin to the previous results on finitary supermartingales, the key aspect of our quantitative variant of the Robbins-Siegmund theorem is that it only requires a finitary variant of the above almost-supermartingale condition:

Definition 3.8. Let $\lambda, \varepsilon > 0$ and $N \in \mathbb{N}$. Let (X_n) be a nonnegative sequence of random variables adapted to (\mathcal{F}_n) . The process is called a λ - ε - N -Robbins-Siegmund process w.r.t. nonnegative sequences of random variables (η_n) , (χ_n) if

$$\mathbb{P}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] > (1 + \chi_n)X_n + \eta_n + \varepsilon) < \lambda$$

for all $n < N$.

We now present our quantitative version of the Robbins-Siegmund theorem as appropriately phrased for our finitary almost-supermartingales in the sense of the above definition. Just as the Robbins-Siegmund theorem is a direct extension of the Doob convergence theorem for nonnegative supermartingales, our quantitative result is a direct extension of Theorem 3.5, where the summability assumptions on (η_n) and (χ_n) are interpreted computationally via a pair of moduli of almost sure boundedness $f, h : (0, \infty) \rightarrow \mathbb{N}$, and in the case of (χ_n) we treat the assumption $\sum_{i=0}^{\infty} \chi_i < +\infty$ in the equivalent form $\prod_{i=0}^{\infty} (1 + \chi_i) < +\infty$. While in this way our theorem bears a resemblance to the main result of [46], we stress once more than for our purposes we require a quantitative Robbins-Siegmund theorem in a different form, with uniform learnability replaced by fluctuation bounds, and crucially the almost supermartingale property given in a finitary form.

Theorem 3.9. Let (X_n) , (η_n) , (χ_n) be sequences of nonnegative random variables adapted to a filtration (\mathcal{F}_n) and let $f, h : (0, \infty) \rightarrow \mathbb{N}$ be functions such that

$$\mathbb{P} \left(\sum_{i=0}^{\infty} \eta_i \geq f(\lambda) \right) < \lambda \text{ and } \mathbb{P} \left(\prod_{i=0}^{\infty} (1 + \chi_i) \geq h(\lambda) \right) < \lambda.$$

Further, let $K > \mathbb{E}[X_0]$ and let μ_n be a modulus of absolute continuity for X_n . Then there are functions $Z_{f,h,K}(\lambda, \varepsilon)$, $e_{f,h,K}(\lambda, \varepsilon, N)$ and $p_{f,h,K}(\lambda, \varepsilon, N)$, that satisfy the following: For any $\lambda, \varepsilon > 0$ and $N \in \mathbb{N}$, if (X_n) is a $(p_{f,h,K}(\lambda, \varepsilon, N))$ - $(e_{f,h,K}(\lambda, \varepsilon, N))$ - N -Robbins-Siegmund process w.r.t. (η_n) , (χ_n) , that is

$$\mathbb{P}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] > (1 + \chi_n)X_n + \eta_n + e_{f,h,K}(\lambda, \varepsilon, N)) < p_{f,h,K}(\lambda, \varepsilon, N)$$

for all $n < N$, then

$$\mathbb{P}(J_{N,\varepsilon}(X_n) \geq Z_{f,h,K}(\lambda, \varepsilon)) < \lambda.$$

Moreover, $Z_{f,h,K}$, $e_{f,h,K}$ and $p_{f,h,K}$ can be explicitly defined via

$$Z_{f,h,K}(\lambda, \varepsilon) := 4 \left\lceil \frac{2^{17}(K + f(\lambda/4) + 1)^2 h(\lambda/10)^2}{\lambda^2 \varepsilon^2} \right\rceil$$

with

$$e_{f,h,K}(\lambda, \varepsilon, N) := \min \left\{ \frac{\varepsilon_0 \lambda_0}{8N}, \frac{\lambda(b/2 + \alpha)}{8}, \frac{\varepsilon}{2N} \right\},$$

$$p_{f,h,K}(\lambda, \varepsilon, N) := \min \left\{ \mu_N^M \left(\frac{\varepsilon_0 \lambda_0}{16N} \right), \mu_N^M \left(\frac{\varepsilon}{8N^2} \right) \right\}.$$

and where $\varepsilon_0 := \varepsilon^2/256Nh(\lambda/10)^2$ and $\lambda_0 := \min\{\mu_e^M(\varepsilon^2/512e^2h(\lambda/10)^2) \mid e \leq N\}$ and, as before, $\mu_N^M(\varepsilon) := \min\{\mu_n(\varepsilon) \mid n \leq N\}$.

Proof. We define the sequences $(Y_n), (Z_n)$ of nonnegative random variables by

$$Y_n := \prod_{i=0}^{n-1} (1 + \chi_i) \quad \text{and} \quad Z_n := \sum_{i=0}^{n-1} \frac{\eta_i}{Y_{i+1}},$$

noting that both are monotone. We then define

$$U_n := \frac{X_n}{Y_n} - Z_n.$$

Through a repeated application of Lemma 3.2, we have for any $a, b, c > 0$ and $N \in \mathbb{N}$:

$$\begin{aligned} \mathbb{P}(J_{N,\varepsilon}(X_i) \geq a) &\leq \mathbb{P}\left(J_{N,\frac{\varepsilon}{c}}(U_i + Z_i) \geq \frac{a}{2}\right) + \mathbb{P}\left(J_{N,\frac{\varepsilon}{b}}(Y_i) \geq \frac{a}{2}\right) \\ &\quad + \mathbb{P}\left(\max_{i \leq N} (|U_i + Z_i|) \geq b\right) + \mathbb{P}\left(\max_{i \leq N} Y_i \geq c\right) \\ &\leq \mathbb{P}\left(J_{N,\frac{\varepsilon}{2c}}(U_i) \geq \frac{a}{4}\right) + \mathbb{P}\left(J_{N,\frac{\varepsilon}{2c}}(Z_i) \geq \frac{a}{4}\right) + \mathbb{P}\left(J_{N,\frac{\varepsilon}{b}}(Y_i) \geq \frac{a}{2}\right) \\ &\quad + \mathbb{P}\left(\max_{i \leq N} |U_i| \geq \frac{b}{2}\right) + \mathbb{P}\left(\max_{i \leq N} Z_i \geq \frac{b}{2}\right) + \mathbb{P}\left(\max_{i \leq N} Y_i \geq c\right). \end{aligned}$$

Now, let $\varepsilon, \lambda > 0$ and $N \in \mathbb{N}$ be given. Our aim is to now define $a, b, c > 0$ and $\mu, \delta > 0$ so that if (X_n) is a μ - δ - N -Robbins-Siegmund process w.r.t. $(\eta_n), (\chi_n)$, then the first term on the right hand side above is bounded by $\lambda/2$ and the remaining by $\lambda/10$, so that combined we have

$$\mathbb{P}(J_{N,\varepsilon}(X_i) \geq a) < \lambda.$$

We first note that since $Y_i \geq 1$ for all $i \in \mathbb{N}$, we get

$$\mathbb{P}\left(\max_{i \leq N} Z_i \geq \frac{b}{2}\right) \leq \mathbb{P}\left(\sup_{i \in \mathbb{N}} Z_i \geq \frac{b}{2}\right) \leq \mathbb{P}\left(\sum_{i=0}^{\infty} \eta_i \geq \frac{b}{2}\right) < \frac{\lambda}{10}$$

for $b \geq 2f\left(\frac{\lambda}{10}\right)$. Similarly, we have

$$\mathbb{P}\left(\max_{i \leq N} Y_i \geq c\right) \leq \mathbb{P}\left(\sup_{i \in \mathbb{N}} Y_i \geq c\right) \leq \mathbb{P}\left(\prod_{i=0}^{\infty} (1 + \chi_i) \geq c\right) < \frac{\lambda}{10}$$

for $c \geq h\left(\frac{\lambda}{10}\right)$. Further, using Lemma 3.7, we have

$$\mathbb{P}\left(J_{N, \frac{\varepsilon}{2c}}(Z_i) \geq \frac{a}{4}\right) < \frac{\lambda}{10} \quad \text{for } a \geq \frac{16c}{\varepsilon} \cdot f\left(\frac{\lambda}{10}\right)$$

and similarly

$$\mathbb{P}\left(J_{N, \frac{\varepsilon}{b}}(Y_i) \geq \frac{a}{2}\right) < \frac{\lambda}{10} \quad \text{for } a \geq \frac{4b}{\varepsilon} \cdot h\left(\frac{\lambda}{10}\right)$$

It therefore remains to bound fluctuations and maxima for (U_n) . Now suppose (X_n) is a μ - δ - N -Robbins-Siegmund process w.r.t. $(\eta_n), (\chi_n)$ for some $\mu, \delta > 0$. Then we must have that (U_n) is a μ - δ - N -supermartingale w.r.t. (\mathcal{F}_n) . To see this, let $n < N$ and let A denote the event

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq (1 + \chi_n)X_n + \eta_n + \delta.$$

Then by standard properties of the conditional expectation, there is a set Ω' of measure one such that on $\Omega' \cap A$ we have

$$\begin{aligned} \mathbb{E}[U_{n+1} | \mathcal{F}_n] &= \mathbb{E}\left[\frac{X_{n+1}}{Y_{n+1}} - Z_{n+1} | \mathcal{F}_n\right] \\ &= \frac{\mathbb{E}[X_{n+1} | \mathcal{F}_n]}{Y_{n+1}} - Z_{n+1} \\ &\leq \frac{(1 + \chi_n)X_n + \eta_n + \delta}{Y_{n+1}} - Z_{n+1} \\ &= \frac{X_n}{Y_n} + \left(\frac{\eta_n}{Y_{n+1}} - Z_{n+1}\right) + \frac{\delta}{Y_{n+1}} \\ &\leq \frac{X_n}{Y_n} - Z_n + \delta \\ &= U_n + \delta. \end{aligned}$$

Therefore, we in particular have

$$\mathbb{P}(\mathbb{E}[U_{n+1} | \mathcal{F}_n] > U_n + \delta) \leq \mathbb{P}(\mathbb{E}[X_{n+1} | \mathcal{F}_n] > (1 + \chi_n)X_n + \eta_n + \delta) < \mu$$

for any such $n < N$. Now, for arbitrary $\alpha > 0$, we define T_α by

$$T_\alpha := \inf\{j \in \mathbb{N} \mid Z_{j+1} > \alpha\}.$$

Since Z_{j+1} is \mathcal{F}_j -measurable for all $j \in \mathbb{N}$, this forms a stopping time. By Lemma 3.3, the stopped process $(U_{i \wedge T_\alpha})$ is also a μ - δ - N -supermartingale, and moreover, since

$$U_{i \wedge T_\alpha} \geq -Z_{i \wedge T_\alpha} \geq -\alpha,$$

the process $(U_{i \wedge T_\alpha} + \alpha)$ is a nonnegative μ - δ - N -supermartingale w.r.t. (\mathcal{F}_n) . Next, we observe that

$$\mathbb{P}\left(J_{N, \frac{\varepsilon}{2c}}(U_i) \geq \frac{a}{4}\right) \leq \mathbb{P}\left(J_{N, \frac{\varepsilon}{2c}}(U_i) \geq \frac{a}{4} \cap \max_{i \leq N} Z_{i+1} < f\left(\frac{\lambda}{4}\right)\right) + \mathbb{P}\left(Z_{N+1} \geq f\left(\frac{\lambda}{4}\right)\right)$$

and since

$$\mathbb{P}\left(Z_{N+1} \geq f\left(\frac{\lambda}{4}\right)\right) \leq \mathbb{P}\left(\sum_{i=0}^{\infty} \eta_i \geq f\left(\frac{\lambda}{4}\right)\right) < \frac{\lambda}{4}$$

it suffices to show that

$$\mathbb{P}\left(J_{N, \frac{\varepsilon}{2c}}(U_i) \geq \frac{a}{4} \cap \max_{i \leq N} Z_{i+1} < f\left(\frac{\lambda}{4}\right)\right) < \frac{\lambda}{4}.$$

For this, we now fix $\alpha := f(\lambda/4)$ so that if ω is in the above measured set, we must have, $T_\alpha(\omega) > N$ and thus $U_{i \wedge T_\alpha}(\omega) = U_i(\omega)$ for all $i \leq N$. So in particular

$$\begin{aligned} \mathbb{P} \left(J_{N, \frac{\varepsilon}{2c}}(U_i) \geq \frac{a}{4} \cap \max_{i \leq N} Z_{i+1} < f \left(\frac{\lambda}{4} \right) \right) &\leq \mathbb{P} \left(J_{N, \frac{\varepsilon}{2c}}(U_{i \wedge T_\alpha}) \geq \frac{a}{4} \right) \\ &= \mathbb{P} \left(J_{N, \frac{\varepsilon}{2c}}(U_{i \wedge T_\alpha} + \alpha) \geq \frac{a}{4} \right), \end{aligned}$$

where the last inequality follows from the fact that the fluctuations of a process are unchanged by adding a constant. We can now appeal to Theorem 3.5, also noting that

$$\mathbb{E}[U_{0 \wedge T_\alpha} + \alpha] = \mathbb{E}[X_0] + \alpha < K + \alpha,$$

by which we have

$$\mathbb{P} \left(J_{N, \frac{\varepsilon}{2c}}(U_{i \wedge T_\alpha} + \alpha) \geq \frac{a}{4} \right) < \frac{\lambda}{4}$$

for $\delta \leq \varepsilon_0 \lambda_0 / 8N$ and $\mu \leq \mu_N^M(\varepsilon_0 \lambda_0 / 16N)$ with $\varepsilon_0 := \varepsilon^2 / 256Nc^2$ and $\lambda_0 := \min\{\mu_e^M(\varepsilon^2 / 512e^2c^2) \mid e \leq N\}$ as well as

$$a \geq 4 \left\lceil \frac{2^{17}(K + \alpha + 1)^2 c^2}{\lambda^2 \varepsilon^2} \right\rceil.$$

Finally, using a similar argument, we see that

$$\begin{aligned} \mathbb{P} \left(\max_{i \leq N} |U_i| \geq \frac{b}{2} \right) &\leq \mathbb{P} \left(\max_{i \leq N} |U_i| \geq \frac{b}{2} \cap \max_{i \leq N} Z_{i+1} < \alpha \right) + \frac{\lambda}{4} \\ &\leq \mathbb{P} \left(\max_{i \leq N} |U_{i \wedge T_\alpha}| \geq \frac{b}{2} \right) + \frac{\lambda}{4} \end{aligned}$$

and if $b/2 > \alpha$, then since $U_{i \wedge T_\alpha} \geq -\alpha$ we have $|U_{i \wedge T_\alpha}| \geq b/2$ if and only if $U_{i \wedge T_\alpha} \geq b/2$, and thus

$$\mathbb{P} \left(\max_{i \leq N} |U_{i \wedge T_\alpha}| \geq \frac{b}{2} \right) = \mathbb{P} \left(\max_{i \leq N} (U_{i \wedge T_\alpha} + \alpha) \geq \frac{b}{2} + \alpha \right) \leq \frac{K + \alpha}{b/2 + \alpha} + \frac{\delta}{b/2 + \alpha}$$

with the final inequality following from Lemma 3.4 if $\mu \leq \mu_N^M(\varepsilon/8N^2)$ and $\delta \leq \varepsilon/2N$. Therefore

$$\mathbb{P} \left(\max_{i \leq n} |U_{i \wedge T_\alpha}| \geq \frac{b}{2} \right) < \frac{\lambda}{4}$$

if $\frac{b}{2} > \alpha$ and $\delta < \frac{\lambda(b/2 + \alpha)}{8}$ and $b \geq \frac{16}{\lambda}(K + \alpha)$. We now put together all of our conditions to obtain

$$\begin{aligned} \alpha &:= f \left(\frac{\lambda}{4} \right), \quad c := h \left(\frac{\lambda}{10} \right), \quad b := \max \left\{ 2f \left(\frac{\lambda}{10} \right), 2\alpha + 1, \frac{16(K + \alpha)}{\lambda} \right\}, \\ a &:= \max \left\{ \frac{16c}{\varepsilon} \cdot f \left(\frac{\lambda}{10} \right), \frac{4b}{\varepsilon} \cdot h \left(\frac{\lambda}{10} \right), 4 \left\lceil \frac{2^{17}(K + \alpha + 1)^2 c^2}{\lambda^2 \varepsilon^2} \right\rceil \right\}, \\ \delta &:= \min \left\{ \frac{\varepsilon_0 \lambda_0}{8N}, \frac{\lambda(b/2 + \alpha)}{8}, \frac{\varepsilon}{2N} \right\}, \\ \mu &:= \min \left\{ \mu_N^M \left(\frac{\varepsilon_0 \lambda_0}{16N} \right), \mu_N^M \left(\frac{\varepsilon}{8N^2} \right) \right\}, \end{aligned}$$

with $\varepsilon_0 := \varepsilon^2 / 256Nc^2$ and $\lambda_0 := \min\{\mu_e^M(\varepsilon^2 / 512e^2c^2) \mid e \leq N\}$ as before. The value a can be simplified to

$$a := 4 \left\lceil \frac{2^{17}(K + f(\lambda/4) + 1)^2 h(\lambda/10)^2}{\lambda^2 \varepsilon^2} \right\rceil.$$

All in all, for these values we have shown that if (X_n) is a μ - δ - N -Robbins-Siegmund process w.r.t. $(\eta_n), (\chi_n)$, then $\mathbb{P}(J_{n,\varepsilon}(X_i) \geq a) < \lambda$ and from this, the result follows. \square

Remark 3.10. Akin to Remarks 2.10 and 3.6, if (X_n) is already a finitary Robbins-Siegmund process in the stronger sense that the associated probability is not resolved, then the above Theorem 3.9 can be simplified in a way that the resulting quantitative information no longer depends on associated moduli of absolute continuity. The proof then in particular relies on the more uniform variant of Theorem 3.5 that is discussed in Remark 3.6.

4. QUANTITATIVE CONVERGENCE OF STOCHASTIC QUASI-FEJÉR MONOTONE SEQUENCES

We now come to the secon main part of the paper, where we combine our quantitative theory of finitary almost-supermartingales with a stochastic lift of the quantitative framework for quasi-Fejér monotone sequences in [35] to provide a quantitative convergence theorem for stochastic quasi-Fejér monotonicity in the presence of compactness. We begin with a series of definitions, most of which are stochastic analogues of simpler concepts from [35].

Throughout, as mentioned before, let (X, d) be a metric space with a fixed but arbitrary point $o \in X$ acting as a center and let $(\Omega, \mathbb{F}, \mathbb{P})$ be a given probability space. On X , let $F \subseteq X$ be a non-empty set. Our main concern is then with sequences of X -valued random variables (x_n) which are stochastically quasi-Fejér monotone relative to F in the following sense, generalizing the seminal notion as e.g. investigated in [17] in the setting of Hilbert spaces⁹ to general metric spaces, suitable to the present setting.

Definition 4.1. Let $\mathcal{F} = (\mathcal{F}_n)$ be a filtration of \mathbb{F} . An X -valued stochastic process (x_n) adapted to \mathcal{F} is stochastically quasi-Fejér monotone w.r.t. F and \mathcal{F} if

$$\forall n \in \mathbb{N} \forall z \in F \left(\mathbb{E}[G(d(x_{n+1}, z)) \mid \mathcal{F}_n] \leq (1 + \chi_n)G(d(x_n, z)) + \eta_n \right) \quad \mathbb{P}\text{-a.s.}$$

where $G : [0, \infty) \rightarrow [0, \infty)$ is an increasing and continuous function with $G(0) = 0$, and $(\eta_n), (\chi_n) \in \ell_+^1(\mathcal{F})$.

These sequences enjoy particularly nice convergence properties if they are combined with a suitable approximation property towards the set F , regarded as an abstract representation of a set of solutions. Following [35], we first encapsulate abstractly what being approximate to being a point in F means here by presuming a decomposition

$$F = \bigcap_{k \in \mathbb{N}} AF_k$$

for a sequence (AF_k) of measurable subsets of X with the property that $AF_{k+1} \subseteq AF_k$ for any $k \in \mathbb{N}$. The set AF_k serves as an abstract representation of the set of all points conceived of as being “ k -good” approximations of being in F . In this abstract formulation, the present setup for handling solution sets and their approximations is indeed very general and immediately encompasses many if not most problems naturally occurring in the literature. Deterministic examples are discussed at length in [35], while a genuine stochastic example will be studied in Section 5 later on.

The approximation property we then consider here is then the following particularly mild assumption, which forms a stochastic analogue of [35, Definition 3.1, (ii)]:

Definition 4.2. An X -valued stochastic process (x_n) has the stochastic lim inf-property w.r.t. F and (AF_k) if

$$\forall k, N \in \mathbb{N} \exists n \geq N (x_n \in AF_k) \quad \mathbb{P}\text{-a.s.}$$

⁹Recall footnotes 4 – 6 from the discussion in the introduction which in detail discuss the motivation for the slight differences between the notion studied in [17] and the one presented in the following.

Both stochastic quasi-Fejér monotonicity and the above approximation property need to be treated in a quantitative way, which we now discuss in detail, beginning with the latter. Under the assumption of these quantitative variants, we will then provide a quantitative almost sure convergence result for stochastic quasi-Fejér monotone sequences under a relative compactness assumption on X .

Notational convention. While we so far have used general real numbers $\varepsilon > 0$ to represent arbitrarily small approximate errors, which is arguably the most natural way, we in this section choose to represent said quantities by expressions of the form $1/(k + 1)$ for $k \in \mathbb{N}$. This serves multiple purposes: For one, it is consistent with earlier papers on the topic of quantitative aspects of Fejér monotone sequences. For another, and arguably more importantly, this representation will allow one to easily judge the measurability of various involved sets in the following parts, which feature various complicated quantitative notions inside probabilities. In order to now freely move between these conventions without onerous bureaucracy, we make use of a slight abuse of notation where for a function $f(\varepsilon)$ taking as argument some real $\varepsilon > 0$, for $k \in \mathbb{N}$ we write $f(k)$ to implicitly mean $f(1/(k + 1))$ (for example, we write $J_{N,k}(X_n)$ for the number of $1/(k + 1)$ -fluctuations experienced by (X_n) before time N). Conversely, if $g(k)$ is a function that refers to the quantity $1/(k + 1)$, for $\varepsilon > 0$ we write $g(\varepsilon)$ to mean $g(k)$ for the least $k \in \mathbb{N}$ with $1/(k + 1) < \varepsilon$. In practice, this notation is completely unambiguous. While we represent approximate arithmetical errors in this way, we do not utilize such a special representation the probabilistic error, usually denoted by λ , in the following, as there the above considerations on e.g. measurability do not apply. While this creates a kind of mixed notation, using natural numbers for arithmetical errors and arbitrary reals for probabilistic errors, we think that it still simplifies the presentation to stick with the “mathematically natural” representation as much as possible.

4.1. Stochastic lim inf-moduli. We start by noting that through elementary properties of probability spaces (recall Lemma 1.1), the stochastic lim inf-property discussed in Definition 4.2 above can be rewritten in such a way that the hidden quantifiers in the almost sure requirement are expanded and then brought outside of the probability measure. Concretely, note that we can equivalently write

$$\begin{aligned} & \forall k, N \in \mathbb{N} \exists n \geq N (x_n \in AF_k) \quad \mathbb{P}\text{-a.s.} \\ & \text{if, and only if, } \mathbb{P}(\forall k, N \in \mathbb{N} \exists n \geq N (x_n \in AF_k)) = 1 \\ & \text{if, and only if, } \forall \lambda > 0 \forall k, N \in \mathbb{N} \exists m \in \mathbb{N} (\mathbb{P}(\forall n \in [N; m] (x_n \notin AF_k)) < \lambda). \end{aligned}$$

Therefore the stochastic lim inf-property for an X -valued stochastic process (x_n) is equivalent to the existence a so-called stochastic lim inf-modulus in the sense of the following definition, providing a bound (and hence a witness) for the quantifier over m in the last line above:

Definition 4.3. A function $\Phi : (0, \infty) \times \mathbb{N}^2 \rightarrow \mathbb{N}$ is a stochastic lim inf-modulus w.r.t. F and (AF_k) for an X -valued stochastic process (x_n) if

$$\forall \lambda > 0 \forall k, N \in \mathbb{N} (\mathbb{P}(\forall n \in [N; \Phi(\lambda, k, N)] (x_n \notin AF_k)) < \lambda).$$

At a crucial point in the course of the proof of the main result, we will however have to rely on a qualitatively equivalent but quantitatively strengthened variant of the above modulus, namely we will require the existence of a function $\Psi : (0, \infty) \times \mathbb{N}^2 \rightarrow \mathbb{N}$ such that

$$(*) \quad \forall \lambda > 0 (\mathbb{P}(\exists k, N \in \mathbb{N} \forall n \in [N; \Psi(\lambda, k, N)] (x_n \notin AF_k)) < \lambda).$$

Naturally such a modulus is also a stochastic lim inf-modulus in the sense of Definition 4.3 and so is a priori stronger. On the other hand, we can construct a modulus satisfying $(*)$ in terms

of a stochastic lim inf-modulus as the following result shows, though this comes at the expense of an exponential speedup in the probability.

Lemma 4.4. *If Φ is a stochastic lim inf-modulus w.r.t. F and (AF_k) for an X -valued stochastic process (x_n) , then $\Psi(\lambda, k, N) := \Phi(\lambda 2^{-(k+N+2)}, k, N)$ satisfies $(*)$.*

Proof. Let $\lambda > 0$. Then we have

$$\begin{aligned} & \mathbb{P}(\exists k, N \in \mathbb{N} \forall n \in [N; \Phi(\lambda 2^{-(k+N+2)}, k, N)](x_n \notin AF_k)) \\ & \leq \sum_{k=0}^{\infty} \sum_{N=0}^{\infty} \mathbb{P}(\forall n \in [N; \Phi(\lambda 2^{-(k+N+2)}, k, N)](x_n \notin AF_k)) \\ & < \sum_{k=0}^{\infty} \sum_{N=0}^{\infty} \lambda 2^{-(k+N+2)} = \lambda \sum_{k=0}^{\infty} 2^{-(k+1)} \sum_{N=0}^{\infty} 2^{-(N+1)} = \lambda. \quad \square \end{aligned}$$

Remark 4.5 (For logicians). Similar to [35], the above notion of a stochastic lim inf-modulus is motivated by the proof-theoretic methodology of proof mining. In particular, general logical metatheorems for proof mining in probability (see again [43] and in particular also [42]) guarantee that such moduli can in fact be extracted from large classes of proofs of the underlying non-quantitative stochastic lim inf-property, provided all (AF_k) can be expressed via an \exists -formula (in the terminology of [43]).

4.2. Moduli of uniform stochastic quasi-Fejér monotonicity. We now arrive at a uniform strengthening of the notion of being stochastically quasi-Fejér monotone, again following the precedent set out in [35] but expanding everything to the stochastic setting. To begin with, we highlight the hidden quantifiers in $z \in F$ in the formulation of the property in Definition 4.1 by expanding it into $\forall k \in \mathbb{N} (z \in AF_k)$:

$$\begin{aligned} & \forall n \in \mathbb{N} \forall z \in X (\forall k \in \mathbb{N} (z \in AF_k) \\ & \rightarrow \forall m \leq n (\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] \leq (1 + \chi_m)G(d(x_m, z)) + \eta_m \quad \mathbb{P}\text{-a.s.})). \end{aligned}$$

Note that we also have introduced an additional bounded quantifier over m , which does not change the property but has the effect that the inner matrix of the property becomes monotone, which is particularly preferable in the context of probabilistic notions. Next, we expand the hidden quantifier in the inequality \leq and in the \mathbb{P} -a.s. property to arrive at the following equivalent phrasing:

$$\begin{aligned} & \forall n \in \mathbb{N} \forall z \in X \left(\forall k \in \mathbb{N} (z \in AF_k) \rightarrow \forall \lambda > 0 \forall r \in \mathbb{N} \forall m \leq n \right. \\ & \left. \left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1} \right) < \lambda \right) \right). \end{aligned}$$

Suitably prenexed, this is in turn equivalent to

$$\begin{aligned} & \forall \lambda > 0 \forall r, n \in \mathbb{N} \forall z \in X \exists k \in \mathbb{N} \forall m \leq n \\ & \left(z \in AF_k \rightarrow \left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1} \right) < \lambda \right) \right). \end{aligned}$$

We now consider a uniform variant of this property, where such a k exists uniformly for all points z , giving rise to a new, stronger notion of stochastic Fejér monotonicity:

Definition 4.6. Let $\mathcal{F} = (\mathcal{F}_n)$ be a filtration of F . An X -valued stochastic process (x_n) adapted to \mathcal{F} is uniformly stochastically quasi-Fejér monotone w.r.t. F , (AF_k) and \mathcal{F} if

$$\forall \lambda > 0 \forall r, n \in \mathbb{N} \exists k \in \mathbb{N} \forall z \in AF_k \forall m \leq n$$

$$\left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1} \right) < \lambda \right)$$

where $G : [0, \infty) \rightarrow [0, \infty)$ is an increasing and continuous function with $G(0) = 0$, and $(\eta_m), (\chi_m) \in \ell_+^1(\mathcal{F})$.

The desired quantitative information that we seek to suitably represent the stochastic quasi-Fejér monotonicity of the process will be a modulus providing a bound (and hence a witness) for for the k in terms of λ and r, n in the above property:

Definition 4.7. A function $\zeta : (0, \infty) \times \mathbb{N}^2 \rightarrow \mathbb{N}$ is a modulus of uniform stochastic quasi-Fejér monotonicity w.r.t. F , (AF_k) and \mathcal{F} for an X -valued stochastic process (x_n) if

$$\forall \lambda > 0 \forall r, n \in \mathbb{N} \forall z \in AF_{\zeta(\lambda, r, n)} \forall m \leq n$$

$$\left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1} \right) < \lambda \right).$$

In fact, it can be shown that this notion coincides with the usual notion of stochastic quasi-Fejér monotonicity over compact spaces, and under suitable continuity assumptions, and we discuss this later in the context of the proof of Theorem 4.22.

Remark 4.8 (For logicians). Similar to [35], the above notion of a modulus of uniform stochastic quasi-Fejér monotonicity is motivated by the proof-theoretic methodology of proof mining and also here, general logical metatheorems for proof mining in probability (we again refer to [43] and [42]) guarantee that such moduli can in fact be extracted from large classes of proofs of the underlying non-quantitative stochastic quasi-Fejér monotonicity property, provided all (AF_k) can be expressed via a \forall -formula (in the terminology of [43]). While this seems to be add odds with the existential condition on the sets (AF_k) as discussed in Remark 4.5, this tension can be resolved in many practical circumstances and we refer to [35] for discussions on that matter.

4.3. A quantitative almost sure convergence theorem under relative compactness.

We now move on to our main quantitative result, providing an effective almost convergence result for uniformly stochastic quasi-Fejér monotone sequences which have the stochastic lim inf-property. As mentioned before, this will take place under a relative compactness assumption which also has to be effectivized:

Definition 4.9 (essentially [28]). For $A \subseteq X$, a function $\gamma : \mathbb{N} \rightarrow \mathbb{N}$ is called a modulus of total boundedness for A if for all $k \in \mathbb{N}$ and any $(x_n) \subseteq A$:

$$\exists 0 \leq i < j \leq \gamma(k) \left(d(x_i, x_j) \leq \frac{1}{k+1} \right).$$

It can be easily seen that a set A permits such a modulus if, and only if, it is totally bounded in the usual sense (see also the discussion in [35]). Also, note that every modulus of total boundedness γ satisfies $\gamma(k) \geq 1$ for any $k \in \mathbb{N}$. For a wide range of examples of totally bounded sets where such moduli can be readily computed, we refer to [35]. For instance, consider a bounded set $A \subseteq H$ with bound b in a finite dimensional Hilbert space H with dimension d . Then

$$\gamma(k) := \left\lceil 2(k+1)\sqrt{db} \right\rceil^d$$

is a modulus of total boundedness for A (cf. Example 2.8 in [35]).

For the rest of this section, unless stated otherwise, we will assume that $AF_k \subseteq X_0$ for any $k \in \mathbb{N}$ where $X_0 \subseteq X$ is a totally bounded set.

We now build up to our main result via a sequence of lemmas, whose role we each briefly outline in advance of their statement and proof.

The first such result is the following Lemma 4.10, which represents the main instantiation of our quantitative Robbins-Siegmund theorem (Theorem 3.9) in the context of uniform stochastic quasi-Fejér monotone sequences, by which we obtain fluctuation bounds for the sequence $(G(x_n, z))$ for all z in a suitably good approximation AF_k to F .

Lemma 4.10. *Let (x_n) be an X -valued stochastic process adapted to a filtration $\mathcal{F} = (\mathcal{F}_n)$ and let $(\eta_n), (\chi_n) \in \ell_+^1(\mathcal{F})$. Assume that we have functions f, g such that*

$$\mathbb{P} \left(\sum_{i=0}^{\infty} \eta_i \geq f(\lambda) \right) < \lambda \text{ and } \mathbb{P} \left(\prod_{i=0}^{\infty} (1 + \chi_i) \geq h(\lambda) \right) < \lambda,$$

along with a modulus of uniform stochastic quasi-Fejér monotonicity $\zeta(\lambda, r, n)$ for (x_n) w.r.t. F , (AF_k) and \mathcal{F} , i.e.

$$\forall \lambda > 0 \forall r, n \in \mathbb{N} \forall z \in AF_{\zeta(\lambda, r, n)} \forall m \leq n$$

$$\left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1} \right) < \lambda \right).$$

Further, let c be a bound relative to o for X_0 , i.e. $d(z, o) \leq c$ for all $z \in X_0$. Finally, assume that b_0 is such that $b_0 > \mathbb{E}[G(d(x_0, o) + c)]$ and that μ_n is a modulus of absolute continuity for $G(d(x_n, z))$, uniformly in $z \in X_0$.

Then there are functions $\chi_0(\lambda, k, N)$ and $\psi_0(\lambda, k)$ such that for any $\lambda > 0$ and $k, N \in \mathbb{N}$:

$$\forall z \in AF_{\chi_0(\lambda, k, N)} (\mathbb{P}(J_{N, k}(G(d(x_n, z))) \geq \psi_0(\lambda, k)) < \lambda).$$

In fact, χ_0 and ψ_0 can simply be defined by $\psi_0(\lambda, k) := Z_{f, g, b_0}(\lambda, k)$ and

$$\chi_0(\lambda, k, N) := \zeta(p_{f, g, b_0}(\lambda, k, N), e_{f, g, b_0}(\lambda, k, N), N)$$

with Z, e and p from Theorem 3.9 (also defined in terms of μ_n).

Proof. As G is increasing, note that we have

$$\mathbb{E}[G(d(x_0, z))] \leq \mathbb{E}[G(d(x_0, o) + c)] < b_0$$

for any $z \in X_0$, hence in particular for any $z \in AF_{\zeta(r, n)}$ for any $r, n \in \mathbb{N}$. The lemma then follows, for any $z \in AF_{\chi_0(\lambda, k, N)}$, directly from Theorem 3.9 with $X_n = G(d(x_n, z))$. \square

Remark 4.11. The above Lemma 4.10, and by that also the later Theorem 4.18, assumes that we are equipped with moduli of absolute continuity μ_n which apply to $G(d(x_n, z))$ uniformly over $z \in X_0$. This uniformity over z could be quite strong at times, but in many situations can still be readily derived. First, in case that $G = \text{Id}$, this assumption reduces to the existence of moduli of absolute continuity for $d(x_n, o)$ as if μ_n is a modulus of absolute continuity for $d(x_n, o)$, then by the triangle inequality and the fact that X_0 is bounded relative to o by c , we have that $\tilde{\mu}_n(\varepsilon) := \min\{\mu_n(\varepsilon/2), \varepsilon/2c\}$ is a modulus of absolute continuity for $d(x_n, z)$, uniformly over $z \in X_0$. For more complicated G , such a modulus might potentially be more difficult to derive in certain situations. However, in most typical situations these are also readily available, already from the way the iterative method (x_n) is defined. A concrete example for such a case, featuring an extremely common type of iteration, is given in Section 5. However, we also want to remind of Remark 2.5 here, by which the existence of such moduli can often

be guaranteed through results from mathematical logic. Lastly, also recall further the previous Remarks 2.10, 3.6 and 3.10 (and note the upcoming Remark 4.20), by which the dependence on such moduli disappears under a stronger approximate stochastic quasi-Fejér monotonicity condition.

Note that whenever $F \neq \emptyset$ and G is coercive, we have that $(d(x_n, o))$ is uniformly almost surely bounded (or finite), i.e.

$$\sup_{n \in \mathbb{N}} d(x_n, o) < \infty \quad \mathbb{P}\text{-a.s.}$$

This follows by applying the Robbins-Siegmund theorem (in its usual, that is non-finitary, form), by which $(G(d(x_n, z)))$ converges and is thus almost surely bounded for any $z \in F$. Since G is coercive, we therefore have that $(d(x_n, z))$, and thus also $(d(x_n, o))$, is almost surely bounded. In the following, we will crucially rely on the assumption that $(d(x_n, o))$ is uniformly almost surely bounded, and hence will require a quantitative rendering of that property. For that purpose, we introduce the notion of a modulus of uniform almost sure boundedness for the random variables $(d(x_n, o))$, that is a function $b : (0, \infty) \rightarrow \mathbb{N}$ such that

$$\forall \lambda > 0 \left(\mathbb{P}(\exists n \in \mathbb{N} (d(x_n, o) > b_\lambda)) < \lambda \right),$$

which will feature crucially in the following results.¹⁰

Remark 4.12. Note that such a modulus of uniform almost sure boundedness for $(d(x_n, o))$ can be derived from an upper bound for the distance $d(z, o)$ for a solution $z \in F$ together with a modulus of coercivity κ for G , i.e. such that κ satisfies $G(x) \geq b$ for all $x \geq \kappa(b)$, by analysing the above argument. Concretely, one can first derive a modulus of uniform almost sure boundedness for $(G(d(x_n, z)))$ via the quantitative Robbins-Siegmund theorem of [46] (see Corollary 3.3 therein), which applies under the same quantitative assumptions on (η_n) and (χ_n) as used in this paper. This can then be converted to a corresponding modulus for $(d(x_n, o))$ using the modulus of coercivity. Instead of resolving this argument fully, that is assuming $F \neq \emptyset$ and the coercivity of G everywhere and dealing with the resulting complex modulus explicitly, we here leave this construction as a sketch, preferring a more abstract approach where we simply assume directly that $(d(x_n, o))$ is uniformly almost surely finite and that we are given a modulus of uniform almost sure boundedness for $d(x_n, o)$.

The following Lemma 4.13 now outlines conditions which allow us to bring the quantification over z inside the measure, controlling the probability that there exists some z such that $(G(x_n, z))$ experiences a large number of fluctuations.

Lemma 4.13. *Let G be uniformly continuous on bounded sets with a modulus $\iota : (0, \infty) \times \mathbb{N} \rightarrow \mathbb{N}$, i.e.*

$$\forall x, y \in [0, b] \left(|x - y| \leq \frac{1}{\iota_b(k) + 1} \rightarrow |G(x) - G(y)| \leq \frac{1}{k + 1} \right)$$

for all $b > 0$ and $k \in \mathbb{N}$. Let $\gamma(k)$ be a modulus of total boundedness, d be a bound relative to o for X_0 , and let (x_n) be a sequence of X -valued random variables with functions χ_0, ψ_0 such that for any $\lambda > 0$ and $k, N \in \mathbb{N}$:

$$\forall z \in AF_{\chi_0(\lambda, k, N)} \left(\mathbb{P}(J_{N, k}(G(d(x_n, z))) \geq \psi_0(\lambda, k)) < \lambda \right).$$

Assume that $b : (0, \infty) \rightarrow \mathbb{N}$ is a modulus of uniform almost sure boundedness for $(d(x_n, o))$, i.e.

$$\forall \lambda > 0 \left(\mathbb{P}(\exists n \in \mathbb{N} (d(x_n, o) > b_\lambda)) < \lambda \right).$$

¹⁰This explicit formulation of almost sure boundedness is related to the notion of *tightness*, see [45, Section 4.1] for a discussion of this point.

Then for any $\lambda > 0$ and $k, N \in \mathbb{N}$ and any finite set $Z_0 \subseteq AF_{\chi_1(\lambda, k, N)}$:

$$\mathbb{P}(\exists z \in Z_0 (J_{N, k}(G(d(x_n, z))) \geq \psi_1(\lambda, k))) < \lambda$$

where

$$\begin{aligned} \chi_1(\lambda, k, N) &= \chi_0 \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2, N \right), \\ \psi_1(\lambda, k) &= \psi_0 \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2 \right). \end{aligned}$$

Proof. Fix $\lambda > 0$ and $k, N \in \mathbb{N}$ and write

$$k_1 := 3k+2 \text{ as well as } \lambda_1 := \frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}.$$

By assumption, we have

$$\forall z \in AF_{\chi_0(\lambda_1, k_1, N)} (\mathbb{P}(J_{N, k_1}(G(d(x_n, z))) \geq \psi_0(\lambda_1, k_1)) < \lambda_1).$$

As γ is a modulus of total boundedness for X_0 , it is also a modulus of total boundedness for $AF_{\chi_0(\lambda_1, k_1, N)}$ and so by [35, Proposition 2.4, (ii)] there are

$$a_1, \dots, a_{\gamma(\iota_{b_{\lambda/2+d}}(3k+2))} \in AF_{\chi_0(\lambda_1, k_1, N)} = AF_{\chi_1(\lambda, k, N)}$$

such that for any $z \in AF_{\chi_1(\lambda, k, N)}$, there exists an $h \in [1; \gamma(\iota_{b_{\lambda/2+d}}(3k+2))]$ such that

$$d(z, a_h) \leq \frac{1}{\iota_{b_{\lambda/2+d}}(3k+2) + 1}.$$

First observe that we have

$$\mathbb{P}(J_{N, k_1}(G(d(x_n, a_h))) \geq \psi_1(\lambda, k)) < \lambda_1 = \frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}$$

for any $h \in [1; \gamma(\iota_{b_{\lambda/2+d}}(3k+2))]$ and so in particular

$$\mathbb{P}(\exists h \in [1; \gamma(\iota_{b_{\lambda/2+d}}(3k+2))]) (J_{N, k_1}(G(d(x_n, a_h))) \geq \psi_1(\lambda, k)) < \frac{\lambda}{2}.$$

Now let ω be such that

$$\forall h \in [1; \gamma(\iota_{b_{\lambda/2+d}}(3k+2))]) (J_{N, k_1}(G(d(x_n(\omega), a_h))) < \psi_1(\lambda, k))$$

as well as $d(x_n(\omega), o) \leq b_{\lambda/2}$ for all $n \in \mathbb{N}$. Let $z \in Z_0$ for some fixed finite set $Z_0 \subseteq AF_{\chi_1(\lambda, k, N)}$. By the above, there exists an $h \in [1; \gamma(\iota_{b_{\lambda/2+d}}(3k+2))]$ such that

$$|d(x_i(\omega), z) - d(x_i(\omega), a_h)| \leq d(z, a_h) \leq \frac{1}{\iota_{b_{\lambda/2+d}}(3k+2) + 1}$$

as well as

$$d(x_i(\omega), z), d(x_i(\omega), a_h) \leq b_{\lambda/2} + d$$

for any $i \in \mathbb{N}$. Thus, we have

$$|G(d(x_i(\omega), z)) - G(d(x_i(\omega), a_h))| \leq \frac{1}{3(k+1)}$$

for any such i and this in particular yields that for any $i, j \in \mathbb{N}$:

$$\begin{aligned} |G(d(x_i(\omega), z)) - G(d(x_j(\omega), z))| &\leq |G(d(x_i(\omega), z)) - G(d(x_i(\omega), a_h))| \\ &\quad + |G(d(x_i(\omega), a_h)) - G(d(x_j(\omega), a_h))| \\ &\quad + |G(d(x_j(\omega), a_h)) - G(d(x_j(\omega), z))| \\ &\leq \frac{2}{3(k+1)} + |G(d(x_i(\omega), a_h)) - G(d(x_j(\omega), a_h))|. \end{aligned}$$

So, any $1/(k+1)$ -fluctuation of $(G(d(x_i(\omega), z)))_{i=0}^N$ is also a $1/(k_1+1) = 1/3(k+1)$ -fluctuation of $(G(d(x_n(\omega), a_h)))_{i=0}^N$. Therefore we have

$$J_{N,k}(G(d(x_n(\omega), z))) \leq J_{N,k_1}(G(d(x_n(\omega), a_h))) < \psi_1(\lambda, k)$$

and so we have shown that

$$\begin{aligned} &\left\{ \omega \mid \forall h \in [1; \gamma(\nu_{b_{\lambda/2}+d}(3k+2))] (J_{N,k_1}(G(d(x_n(\omega), a_h))) < \psi_1(\lambda, k)) \right\} \\ &\quad \cap \{ \omega \mid \forall n \in \mathbb{N} (d(x_n(\omega), o) \leq b_{\lambda/2}) \} \\ &\subseteq \{ \omega \mid \forall z \in Z_0 (J_{N,k}(G(d(x_n(\omega), z))) < \psi_1(\lambda, k)) \} \end{aligned}$$

and the result follows immediately. \square

We introduce a new kind of ‘set-fluctuation’ parameterised by a finite set Z_0 : Let

$$J_k(G(d(x_n, Z_0)))$$

denote the largest p such that there exist indices $i_1 < j_1 \leq i_2 < j_2 \leq \dots < i_p < j_p$ such that

$$\forall l \leq p \exists z \in Z_0 \left(|G(d(x_{i_l}, z)) - G(d(x_{j_l}, z))| \geq \frac{1}{k+1} \right)$$

and $J_{N,k}(G(d(x_n, Z_0)))$ for the same notion on $(G(d(x_n, Z_0)))_{n=0}^N$.

The next Lemma 4.14 now converts the previous assertion from Lemma 4.13 to a bound on the ‘set-fluctuations’ of the sequence $(G(x_n, Z_0))$ where Z_0 is a finite set in a suitably good approximation AF_j to F .

Lemma 4.14. *Fix $N \in \mathbb{N}$ and let Z_0 be a finite set of size p . Suppose ψ is such that for all $\lambda > 0$ and $k \in \mathbb{N}$:*

$$\mathbb{P}(\exists z \in Z_0 (J_{N,k}(G(d(x_n, z))) \geq \psi(\lambda, k))) < \lambda.$$

Then for all $\lambda > 0$ and $k \in \mathbb{N}$:

$$\mathbb{P}(J_{N,k}(G(d(x_n, Z_0))) \geq p \cdot \psi(\lambda, k)) < \lambda.$$

Proof. Let ω be such that for all $z \in Z_0$:

$$J_{N,k}(G(d(x_n(\omega), z))) < \psi(\lambda, k).$$

By definition, each set- $1/(k+1)$ -fluctuation of $(G(d(x_n(\omega), Z_0)))_{n=0}^N$ represents a standard $1/(k+1)$ -fluctuation of $(G(d(x_n(\omega), z)))_{n=0}^N$ for at least one $z \in Z_0$, and therefore, as by the above there are $< \psi(\lambda, k)$ of such fluctuations for every $z \in Z_0$, there can be only $< p \cdot \psi(\lambda, k)$ of the former set-fluctuations as Z_0 has size p . That is, we have shown that

$$\{ \omega \mid \forall z \in Z_0 (J_{N,k}(G(d(x_n(\omega), z))) < \psi(\lambda, k)) \} \subseteq \{ \omega \mid J_{N,k}(G(d(x_n(\omega), Z_0))) < p \cdot \psi(\lambda, k) \}$$

from which the result follows. \square

The following Lemma 4.15 then converts this set-fluctuation bound from the previous Lemma 4.14 to a corresponding rate of pointwise metastability, guaranteeing that for any $k \in \mathbb{N}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ we can find an interval $[n; n+g(n)]$ and $z \in Z_0$ such that $G(d(x_i, z))$ is $1/(k+1)$ -stable with high probability.

Lemma 4.15. *Let Z_0 be a finite set and fix $\lambda > 0$ as well as $k \in \mathbb{N}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$. Suppose that ψ is such that*

$$\mathbb{P}(J_{N_g^{\lambda,k}}(G(d(x_n, Z_0))) \geq \psi(\lambda, k)) < \lambda.$$

Then also

$$\mathbb{P}\left(\forall n \leq \Psi(\lambda, k, g) \exists i, j \in [n; n+g(n)] \exists z \in Z_0 \left(|G(d(x_i, z)) - G(d(x_j, z))| > \frac{1}{k+1}\right)\right) < \lambda$$

where we define

$$\Psi(\lambda, k, g) := N_g^{\lambda,k} := \tilde{g}^{(\psi(\lambda,k))}(0)$$

with $\tilde{g}(n) := n + g(n)$.

Proof. Let ω be such that

$$\forall n \leq \Psi(\lambda, k, g) \exists i, j \in [n; n+g(n)] \exists z \in Z_0 \left(|G(d(x_i(\omega), z)) - G(d(x_j(\omega), z))| > \frac{1}{k+1}\right).$$

Take $n_0 := 0$ and $n_{l+1} := \tilde{g}(n_l)$ for $l < \psi(\lambda, k)$. Since then $n_l < \Psi(\lambda, k, g)$ for $l < \psi(\lambda, k)$, for each such l there exist pairs $i, j \in [n_l; n_{l+1}]$ together with an accompanying $z \in Z_0$ such that

$$|G(d(x_i(\omega), z)) - G(d(x_j(\omega), z))| > \frac{1}{k+1}.$$

As all these $\psi(\lambda, k)$ -many pairs i, j lie below $N_g^{\lambda,k} = \Psi(\lambda, k, g)$, we have shown

$$J_{N_g^{\lambda,k}}(G(d(x_n(\omega), Z_0))) \geq \psi(\lambda, k),$$

and the result follows immediately. \square

The next result, Lemma 4.16, is now the last one in our line of preparatory results for the main theorem in which we apply a quantitative compactness argument along with the lim inf-property and continuity of G to transform rates of metastability for $G(d(x_i, z))$ in the above sense to a rate of metastability for (x_i) in the usual sense.

Lemma 4.16. *Let G be uniformly continuous on bounded sets with a modulus $\iota : (0, \infty) \times \mathbb{N} \rightarrow \mathbb{N}$, i.e.*

$$\forall x, y \in [0, b] \left(|x - y| \leq \frac{1}{\iota_b(k) + 1} \rightarrow |G(x) - G(y)| \leq \frac{1}{k+1}\right)$$

for all $b > 0$ and $k \in \mathbb{N}$. Also, let G be inverse continuous at 0 with a modulus $\nu : \mathbb{N} \rightarrow \mathbb{N}$, i.e.

$$\forall x \geq 0 \left(G(x) \leq \frac{1}{\nu(k) + 1} \rightarrow x \leq \frac{1}{k+1}\right)$$

for all $k \in \mathbb{N}$. Let $\gamma(k)$ be a modulus of total boundedness and d be a bound relative to o for X_0 and let (x_n) be a sequence of X -valued random variables. Assume that $b : (0, \infty) \rightarrow \mathbb{N}$ is a modulus of uniform almost sure boundedness for $(d(x_n, o))$, i.e.

$$\forall \lambda > 0 (\mathbb{P}(\exists n \in \mathbb{N} (d(x_n, o) > b_\lambda)) < \lambda).$$

Assume that there are Ψ, Z such that for all $\lambda > 0$ and all $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$ as well as $p \in \mathbb{N}$:

$$(I) \quad \forall Z_0 \subseteq AF_{Z(\lambda, k, g, p)} \text{ finite of size } p$$

$$\left(\mathbb{P} \left(\forall m \leq \Psi(\lambda, k, g, p) \exists i, j \in [m; m + g(m)] \exists z \in Z_0 \right. \right. \\ \left. \left. \left(|G(d(x_i, z)) - G(d(x_j, z))| > \frac{1}{k+1} \right) \right) < \lambda \right).$$

Assume further $Z(\lambda, k, g, p) \geq k$. Finally, let Φ be such that for all $\lambda > 0$:

$$(II) \quad \mathbb{P}(\exists k, N \in \mathbb{N} \forall n \in [N; \Phi(\lambda, k, N)] (x_n \notin AF_k)) < \lambda.$$

Then (x_n) is almost surely Cauchy and furthermore, for all $\lambda > 0$ and all $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$:

$$(III) \quad \mathbb{P} \left(\forall n \leq \Delta(\lambda, k, g) \exists i, j \in [n; n + g(n)] \left(d(x_i, x_j) > \frac{1}{k+1} \right) \right) < \lambda,$$

where Δ is defined by

$$\Delta(\lambda, k, g) := \max\{\Delta_i(\lambda, k, g) \mid i \in [0; \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))]\},$$

with

$$\Delta_i(\lambda, k, g) := \max\{\Phi(\lambda/3, k_i^\lambda, m) \mid m \leq \Psi(\widehat{\lambda}_k, k_i^\lambda, h_{k_i^\lambda}^{\lambda/3}, p_k^\lambda)\},$$

with

$$h_l^\mu(m) := \max\{m' + g(m') \mid m' \in [m; \Phi(\mu, l, m)]\} - m,$$

and

$$\begin{cases} k_0^\lambda := \max\{6\nu(2k+1) + 5, k\}, \\ k_{j+1}^\lambda := \max\{Z(\widehat{\lambda}_k, k_i^\lambda, h_{k_i^\lambda}^{\lambda/3}, p_k^\lambda) \mid i \leq j\}, \end{cases}$$

as well as $p_k^\lambda := \gamma(\iota_{\max\{2b_{\lambda/3}, b_{\lambda/3}+d\}}(6\nu(2k+1)+5))$ and

$$\widehat{\lambda}_k := \frac{\lambda}{3(\gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1)) + 1)}.$$

Proof. Given $\lambda, \mu > 0$ and $k, j \in \mathbb{N}$ as well as $g : \mathbb{N} \rightarrow \mathbb{N}$, we denote by

$$A_{\mu, j}^{\lambda, k, g} = \{a_{\mu, j, r}^{\lambda, k, g} \mid r \in [1; \gamma(\iota_{\max\{2b_{\mu/3}, b_{\mu/3}+d\}}(6\nu(2j+1)+5))]\}$$

the $1/(\iota_{\max\{2b_{\mu/3}, b_{\mu/3}+d\}}(6\nu(2j+1)+5)+1)$ -net of $AF_{Z(\lambda, k, g, p_j^\mu)}$ with size

$$p_j^\mu := \gamma(\iota_{\max\{2b_{\mu/3}, b_{\mu/3}+d\}}(6\nu(2j+1)+5))$$

which exists by [35, Proposition 2.4, (ii)] as X_0 , and hence also $AF_{Z(\lambda, k, g, p_j^\mu)}$, is totally bounded with a modulus of total boundedness γ . In other words, the $a_{\mu, j, h}^{\lambda, k, g}$ have the property that

$$\forall z \in AF_{Z(\lambda, k, g, p_j^\mu)} \exists r \in [1; p_j^\mu] \left(d(z, a_{\mu, j, r}^{\lambda, k, g}) \leq \frac{1}{\iota_{\max\{2b_{\mu/3}, b_{\mu/3}+d\}}(6\nu(2j+1)+5)+1} \right).$$

Then, by assumption (I), we get for any $\lambda, \mu > 0$ and $k, j \in \mathbb{N}$ as well as $g : \mathbb{N} \rightarrow \mathbb{N}$:

$$(I') \quad \mathbb{P} \left(\forall m \leq \Psi(\lambda, k, g, p_j^\mu) \exists i, j \in [m; m + g(m)] \right. \\ \left. \exists r \in [1; p_j^\mu] \left(\left| G(d(x_i, a_{\mu, j, r}^{\lambda, k, g})) - G(d(x_j, a_{\mu, j, r}^{\lambda, k, g})) \right| > \frac{1}{k+1} \right) \right) < \lambda.$$

Denote the set measured in (I') by $P_{\lambda, k, g, \mu, j}$ and denote by Q_λ the set measured in (II) and lastly by $R_{\lambda, k, g}$ the set measured in (III). Denote by S_λ the set measured in the assumption of

uniform almost sure boundedness. Fix $\lambda > 0$ and $k \in \mathbb{N}$ as well as $g : \mathbb{N} \rightarrow \mathbb{N}$, and take ω such that

$$\omega \in S_{\lambda/3}^c \cap Q_{\lambda/3}^c \cap \bigcap_{l=0}^{\gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))} P_{\widehat{\lambda}_k, k_l^\lambda, h_{k_l^\lambda}^{\lambda/3}, \lambda, k}^c.$$

Then, by the last part of the intersection, we have

$$(+)_{l_i} \quad \exists m_l \leq \Psi(\widehat{\lambda}_k, k_l^\lambda, h_{k_l^\lambda}^{\lambda/3}, p_k^\lambda) \quad \forall i, j \in [m_l; m_l + h_{k_l^\lambda}^{\lambda/3}(m_l)] \quad \forall r \in [1; p_k^\lambda]$$

$$\left(|G(d(x_i(\omega), a_r^l)) - G(d(x_j(\omega), a_r^l))| \leq \frac{1}{6(\nu(2k+1)+1)} \right)$$

for any $l \in [0; \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))]$, where we write

$$a_r^l = a_{\lambda, k, r}^{\widehat{\lambda}_k, k_l^\lambda, h_{k_l^\lambda}^{\lambda/3}}.$$

and the inner formula follows from the assumed lower bound on Z , specifically that for $l > 0$ we have

$$k_l^\lambda \geq Z(\widehat{\lambda}_k, k_0^\lambda, h_{k_0^\lambda}^{\lambda/3}, p_k^\lambda) \geq k_0^\lambda \geq 6\nu(2k+1)+5,$$

which implies $k_l^\lambda + 1 \geq 6(\nu(2k+1)+1)$ for all such l . Further, by the second part of the intersection, we have

$$\forall k', N' \in \mathbb{N} \quad \exists n' \in [N'; \Phi(\lambda/3, k', N')] \quad (x_{n'}(\omega) \in AF_{k'}).$$

Lastly, by the first part of the intersection, we have $d(x_i(\omega), o) \leq b_{\lambda/3}$ for any $i \in \mathbb{N}$. Now, taking for each $l \in [0; \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))]$ an m_l satisfying $(+)_{l_i}$, we get

$$\exists n_l \in [m_l; \Phi(\lambda/3, k_l^\lambda, m_l)] \quad (x_{n_l}(\omega) \in AF_{k_l^\lambda}).$$

By definition of $h_{k_l^\lambda}^{\lambda/3}$, we have

$$[n_l; n_l + g(n_l)] \subseteq [m_l; m_l + h_{k_l^\lambda}^{\lambda/3}(m_l)]$$

and so we in particular have

$$\exists n_l \leq \Delta_l(\lambda, k, g) \left(x_{n_l}(\omega) \in AF_{k_l^\lambda} \wedge \forall i, j \in [n_l; n_l + g(n_l)] \quad \forall r \in [1; p_k] \right.$$

$$\left. \left(|G(d(x_i(\omega), a_r^l)) - G(d(x_j(\omega), a_r^l))| \leq \frac{1}{6(\nu(2k+1)+1)} \right) \right).$$

for all $l \in [0; \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))]$. Note that $x_{n_l}(\omega) \in X_0$ for any $l \in [0; \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))]$ so that, by using the modulus of total boundedness γ of X_0 on the sequence $(x_{n_l}(\omega))$, we have

$$\exists 0 \leq I < J \leq \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1)) \left(d(x_{n_I}(\omega), x_{n_J}(\omega)) \leq \frac{1}{\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1)+1} \right).$$

For that pair, we thus have

$$G(d(x_{n_I}(\omega), x_{n_J}(\omega))) \leq \frac{1}{2(\nu(2k+1)+1)}.$$

and in particular, by the above, we have

$$x_{n_J}(\omega) \in AF_{k_J}^\lambda \subseteq AF_{Z(\widehat{\lambda}_k, k_I^\lambda, h_{k_I^\lambda}^{\lambda/3}, p_k^\lambda)}$$

as well as

$$\forall r \in [1; p_k^\lambda] \forall i, j \in [n_I; n_I + g(n_I)] \left(|G(d(x_i(\omega), a_r^I)) - G(d(x_j(\omega), a_r^I))| \leq \frac{1}{6(\nu(2k+1)+1)} \right).$$

Therefore, by the properties of the a_r^I 's, we have that there exists an $r \in [1; p_k^\lambda]$ such that

$$|d(x_{n_J}(\omega), x_i(\omega)) - d(a_r^I, x_i(\omega))| \leq d(x_{n_J}(\omega), a_r^I) \leq \frac{1}{\iota_{\max\{2b_{\lambda/3}, b_{\lambda/3+d}\}}(6\nu(2k+1)+5)+1}$$

and so

$$|G(d(x_{n_J}(\omega), x_i(\omega))) - G(d(a_r^I, x_i(\omega)))| \leq \frac{1}{6(\nu(2k+1)+1)}$$

for any $i \in [n_I; n_I + g(n_I)]$. Combined, we have for all $i, j \in [n_I; n_I + g(n_I)]$:

$$|G(d(x_i(\omega), x_{n_J}(\omega))) - G(d(x_j(\omega), x_{n_J}(\omega)))| \leq \frac{1}{2(\nu(2k+1)+1)}.$$

This yields

$$G(d(x_i(\omega), x_{n_J}(\omega))) \leq \frac{1}{\nu(2k+1)+1}$$

and so in particular

$$d(x_i(\omega), x_{n_J}(\omega)) \leq \frac{1}{2(k+1)}$$

for all $i \in [n_I; n_I + g(n_I)]$ from which we can deduce that

$$\forall i, j \in [n_I; n_I + g(n_I)] \left(d(x_i(\omega), x_j(\omega)) \leq \frac{1}{k+1} \right)$$

and so, as $n_I \leq \Delta(\lambda, k, g)$, we get $\omega \in R_{\lambda, k, g}^c$. In summary, since ω was arbitrary, we have shown

$$S_{\lambda/3}^c \cap Q_{\lambda/3}^c \cap \bigcap_{l=0}^{\gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))} P_{\widehat{\lambda}_k, k_l^\lambda, h_{k_l^\lambda}^{\lambda/3}, \lambda, k}^c \subseteq R_{\lambda, k, g}^c$$

and so

$$\begin{aligned} \mathbb{P}(R_{\lambda, k, g}) &\leq \mathbb{P}(S_{\lambda/3}) + \mathbb{P}(Q_{\lambda/3}) + \sum_{l=0}^{\gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))} \mathbb{P}\left(P_{\widehat{\lambda}_k, k_l^\lambda, h_{k_l^\lambda}^{\lambda/3}, \lambda, k}\right) \\ &< \lambda/3 + \lambda/3 + (\gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1)) + 1)\widehat{\lambda}_k = \lambda, \end{aligned}$$

the last step following by definition of $\widehat{\lambda}_k$. \square

Remark 4.17. Lemmas 4.13 – 4.16, if considered deterministically, in fact provide a quantitative version of Opial's lemma (see e.g. Fact 1.1 in the recent [3] and also the surrounding discussion therein) on totally bounded metric spaces in the case where we have bounds on the fluctuations for the key assumption of Opial's lemma that $d(x_k, z)$ converges for all solutions z in a suitably uniform way. This illustrates the considerably combinatorial complexity inherent in Opial's lemma (and in fact already in such special cases). In particular, the fact that we have to resolve this fragment of Opial's lemma in this way is a key reason for the complexity of our final bound and a key difference to the previous quantitative results for deterministic quasi-Fejér monotone sequences given in [35].

We can now derive our main quantitative result. For that, we just recall the setup: We are working over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an arbitrary metric space X , together with a problem set $F = \bigcap_{k \in \mathbb{N}} AF_k$ for a sequence (AF_k) with $AF_{k+1} \subseteq AF_k \subseteq X_0$ for all $k \in \mathbb{N}$, all included in a set $X_0 \subseteq X$ which is totally bounded.

Theorem 4.18. *Let G be uniformly continuous on bounded sets with a modulus $\iota : (0, \infty) \times \mathbb{N} \rightarrow \mathbb{N}$, i.e.*

$$\forall x, y \in [0, b] \left(|x - y| \leq \frac{1}{\iota_b(k) + 1} \rightarrow |G(x) - G(y)| \leq \frac{1}{k + 1} \right)$$

for all $b > 0$ and $k \in \mathbb{N}$. Also, let G be inverse continuous at 0 with a modulus $\nu : \mathbb{N} \rightarrow \mathbb{N}$, i.e.

$$\forall x \geq 0 \left(G(x) \leq \frac{1}{\nu(k) + 1} \rightarrow x \leq \frac{1}{k + 1} \right)$$

for all $k \in \mathbb{N}$. Let $\gamma(k)$ be a modulus of total boundedness and d be a bound relative to o for X_0 . Let (x_n) be a sequence of X -valued random variables adapted to $\mathcal{F} = (\mathcal{F}_n)$ and $(\eta_n), (\chi_n) \in \ell_+^1(\mathcal{F})$. Assume that we have functions f, g such that

$$\mathbb{P} \left(\sum_{i=0}^{\infty} \eta_i \geq f(\lambda) \right) < \lambda \text{ and } \mathbb{P} \left(\prod_{i=0}^{\infty} (1 + \chi_i) \geq h(\lambda) \right) < \lambda.$$

Further assume that we have a modulus of uniform stochastic quasi-Fejér monotonicity $\zeta(r, n)$ for (x_n) w.r.t. F , (AF_k) and \mathcal{F} , i.e.

$$\forall \lambda > 0 \forall r, n \in \mathbb{N} \forall z \in AF_{\zeta(\lambda, r, n)} \forall m \leq n$$

$$\left(\mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r + 1} \right) < \lambda \right).$$

and that (x_n) has the stochastic lim inf-property w.r.t. F and (AF_k) with a corresponding modulus $\Phi(\lambda, k, N)$, i.e.

$$\forall \lambda > 0 \forall k, N \in \mathbb{N} (\mathbb{P}(\forall n \in [N; \Phi(\lambda, k, N)](x_n \notin AF_k)) < \lambda).$$

Assume that $b : (0, \infty) \rightarrow \mathbb{N}$ is a modulus of uniform almost sure boundedness for $(d(x_n, o))$, i.e.

$$\forall \lambda > 0 (\mathbb{P}(\exists n \in \mathbb{N} (d(x_n, o) > b_\lambda)) < \lambda).$$

Lastly, assume that b_0 is such that $b_0 > \mathbb{E}[G(d(x_0, o) + d)]$ and that μ_n is a modulus of absolute continuity for $G(d(x_n, z))$, uniformly in $z \in X_0$.

Then (x_n) is almost surely Cauchy and further, for all $\lambda > 0$ and all $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$:

$$\mathbb{P} \left(\forall n \leq \Delta(\lambda, k, g) \exists i, j \in [n; n + g(n)] \left(d(x_i, x_j) > \frac{1}{k + 1} \right) \right) < \lambda,$$

where Δ is defined by

$$\Delta(\lambda, k, g) := \max\{\Delta_i(\lambda, k, g) \mid i \in [0; \gamma(\iota_{2b_{\lambda/3}}(2\nu(2k + 1) + 1))]\},$$

with

$$\Delta_i(\lambda, k, g) := \max\{\Phi'(\lambda/3, k_i^\lambda, m) \mid m \leq \Psi(\widehat{\lambda}_k, k_i^\lambda, h_{k_i^\lambda}^{\lambda/3}, p_k^\lambda)\},$$

with

$$h_i^\mu(m) := \max\{m' + g(m') \mid m' \in [m; \Phi'(\mu, l, m)]\} - m,$$

and

$$\begin{cases} k_0^\lambda := \max\{6\nu(2k + 1) + 5, k\}, \\ k_{j+1}^\lambda := \max\{\tilde{Z}(\widehat{\lambda}_k, k_i^\lambda, h_{k_i^\lambda}^{\lambda/3}, p_k^\lambda), k_i^\lambda \mid i \leq j\}, \end{cases}$$

as well as $p_k^\lambda := \gamma(\iota_{\max\{2b_{\lambda/3}, b_{\lambda/3+d}\}}(6\nu(2k+1)+5))$ and

$$\widehat{\lambda}_k := \frac{\lambda}{3(\gamma(\iota_{2b_{\lambda/3}}(2\nu(2k+1)+1))+1)},$$

where now Ψ and \tilde{Z} are defined by

$$\begin{aligned} \tilde{Z}(\lambda, k, g, q) := & \zeta \left(p_{f,g,b_0} \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2, N_{g,q}^{\lambda,k} \right), \right. \\ & \left. e_{f,g,b_0} \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2, N_{g,q}^{\lambda,k} \right), N_{g,q}^{\lambda,k} \right) \end{aligned}$$

and

$$\Psi(\lambda, k, g, q) := N_{g,q}^{\lambda,k} := \tilde{g} \left(q^{Z_{f,g,b_0}} \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2 \right) \right)_{(0)}$$

for $\tilde{g}(n) := n + g(n)$ and with Z , e and p from Theorem 3.9 (also defined in terms of μ_n), and Φ' is defined by

$$\Phi'(\lambda, k, N) := \Phi(\lambda 2^{-(k+N+2)}, k, N).$$

Proof. First, combining Lemmas 4.10 and 4.13, we have that for any $\lambda > 0$ and $k, N \in \mathbb{N}$, for any finite set $Z_0 \subseteq AF_{\chi_1(\lambda, k, N)}$

$$\mathbb{P}(\exists z \in Z_0 (J_{N,k}(G(d(x_n, z))) \geq \psi_1(\lambda, k))) < \lambda$$

for

$$\begin{aligned} \chi_1(\lambda, k, N) := & \zeta \left(p_{f,g,b_0} \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2, N \right), \right. \\ & \left. e_{f,g,b_0} \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2, N \right), N \right) \end{aligned}$$

and

$$\psi_1(\lambda, k) := Z_{f,g,b_0} \left(\frac{\lambda}{2\gamma(\iota_{b_{\lambda/2+d}}(3k+2))}, 3k+2 \right).$$

Thus by Lemma 4.14, whenever Z_0 has size q then

$$\mathbb{P}(J_{N,k}(G(d(x_n, Z_0))) \geq q \cdot \psi_1(\lambda, k)) < \lambda,$$

and so by Lemma 4.15 we derive that Ψ and \tilde{Z} defined as above satisfy for all $\lambda > 0$ and all $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$ as well as $q \in \mathbb{N}$:

$$\begin{aligned} & \forall Z_0 \subseteq AF_{\max\{\tilde{Z}(\lambda, k, g, q), k\}} \text{ finite of size } q \\ & \left(\mathbb{P} \left(\forall m \leq \Psi(\lambda, k, g, q) \exists i, j \in [m; m+g(m)] \exists z \in Z_0 \right. \right. \\ & \quad \left. \left. \left(|G(d(x_i, z)) - G(d(x_j, z))| > \frac{1}{k+1} \right) \right) < \lambda \right). \end{aligned}$$

Briefly observing, using Lemma 4.4, that Φ' satisfies

$$\forall \lambda > 0 (\mathbb{P}(\exists k, N \in \mathbb{N} \forall n \in [N; \Phi'(\lambda, k, N)](x_n \notin AF_k)) < \lambda),$$

the result then follows from Lemma 4.16. \square

Remark 4.19. Broadly speaking, the rate of metastability in Theorem 4.18 represents the synthesis of two main components:

- (1) The quantitative Robbins-Siegmund theorem (cf. Theorem 3.9), as brought into the framework of stochastic quasi-Fejér monotonicity by Lemma 4.10 to yield a modulus of finite fluctuations for $G(d(x_n, z))$ for $z \in AF_k$ for k sufficiently large,
- (2) A quantitative version of (the stochastic lifting of) Opial's lemma (cf. Lemmas 4.13 – 4.16 and recall also Remark 4.17) which converts this modulus of finite fluctuations to a pointwise rate of metastability for (x_n) .

While the former carries a greater *conceptual* complexity, and indeed occupied us for the first half of the paper, the resulting bounds are relatively simple. In particular, the fluctuation bound $\psi_0(\lambda, k)$ for $G(d(x_n, z))$ as in Lemma 4.10 is essentially of the form $C_\lambda/\lambda^2\varepsilon^2$, where in the special case that $(\eta_n), (\chi_n)$ are deterministic, $C_\lambda := C$ for some fixed constant, which in turn aligns with the (optimal) bounds for uniform convergence of L_1 -martingales established in [45] (see in particular the discussion in Section 7.4 of that paper). The bound $\chi_0(\lambda, k, N)$ in Lemma 4.10 is new in this paper, but is still nevertheless built up of simple polynomial terms which are taken as arguments by the various moduli ζ, f, g, μ_n .

The second component, on the other hand, carries greater *computational* complexity, notably in the recursive construction in Lemma 4.16. Indeed, it is in Lemma 4.16 that simple fluctuation bounds no longer seem to be possible, forcing us to move to the more general metastable rate of pointwise convergence. Thus, as already highlighted, the main combinatorial complexity of our construction lies in Opial's lemma rather than supermartingale convergence, where the latter, though mathematically more involved, admits bounds that are of a lower complexity and more uniform in nature.

Remark 4.20. Akin to Remarks 2.10, 3.6 and 3.10, if (x_n) is already uniformly quasi-Fejér monotone in the stronger sense that the associated probability is not resolved, then the above Theorem 4.18 can be simplified in a way that the resulting quantitative information no longer depends on associated moduli of absolute continuity. The proof then in particular relies on the more uniform variant of Theorem 3.9 that is discussed in Remark 3.10.

Forgetting about the quantitative data, we obtain the following “plain” almost sure convergence result as an immediate corollary of our main theorem above:

Theorem 4.21. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a metric space. Further, let $X_0 \subseteq X$ be totally bounded and let $F = \bigcap_{k \in \mathbb{N}} AF_k$ for a sequence (AF_k) with $AF_{k+1} \subseteq AF_k \subseteq X_0$ for all $k \in \mathbb{N}$. Let G be increasing and continuous with $G(0) = 0$. Let (x_n) be a sequence of X -valued random variables which is uniformly stochastically quasi-Fejér monotone w.r.t. F , (AF_k) and \mathcal{F} and which has the stochastic liminf-property w.r.t. F and (AF_k) . Assume that $(d(x_n, o))$ is uniformly almost surely bounded and that $G(d(x_0, o) + d)$ has finite mean. Then (x_n) is almost surely Cauchy.*

If we now assume that X is complete and separable, then we can further guarantee that (x_n) converges almost surely to a limit. To guarantee the feasibility of the solution in our abstract setup, we further require a suitable closedness property for F relative to the (AF_k) . For that, we further follow the approach of [35] and consider the abstract property that F is explicitly closed w.r.t. (AF_k) , in the sense that

$$\forall p \in X \ (\forall N, M \in \mathbb{N} \ (AF_M \cap \overline{B}_{1/(N+1)}(p) \neq \emptyset) \rightarrow p \in F).$$

Under this assumption, we can then guarantee that the above limit is almost surely a solution, i.e. that it belongs to F , and hence derive the following extended result. In it, we also further

reduce the property of uniform stochastic quasi-Fejér monotonicity to its ordinary counterpart, and following the discussion before Remark 4.12 replace the assumption that $(d(x_n, o))$ is almost surely finite by the assumptions that $F \neq \emptyset$ and that G is coercive.

Theorem 4.22. *Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space and let X be a complete and separable metric space. Further, let $X_0 \subseteq X$ be compact and let $\emptyset \neq F = \bigcap_{k \in \mathbb{N}} AF_k$ for a sequence (AF_k) with $AF_{k+1} \subseteq AF_k \subseteq X_0$ for all $k \in \mathbb{N}$, such that F is explicitly closed w.r.t. (AF_k) . Let G be increasing, continuous and coercive with $G(0) = 0$. Let (x_n) be a sequence of X -valued random variables which is stochastically quasi-Fejér monotone w.r.t. F , (AF_k) and a filtration \mathcal{F} of \mathbb{F} , and which has the stochastic lim inf-property w.r.t. F and (AF_k) . Assume that $G(d(x_0, o) + d)$ has finite mean. Then $x_n \rightarrow x$ almost surely, where $x \in F$ almost surely.*

Proof. We first show that, in the context of the present assumptions of compactness and explicit closedness, the sequence (x_n) is in fact already uniformly stochastically quasi-Fejér monotone. Suppose for a contradiction that this is not the case, i.e. that there are $\lambda > 0$ and $r, n \in \mathbb{N}$ such that for any $k \in \mathbb{N}$, there is a $z_k \in AF_k$ and an $m_k \leq n$ such that

$$\mathbb{P} \left(\mathbb{E}[G(d(x_{m_k+1}, z_k)) \mid \mathcal{F}_{m_k}] > (1 + \chi_{m_k})G(d(x_{m_k}, z_k)) + \eta_{m_k} + \frac{1}{r+1} \right) \geq \lambda.$$

Using the compactness of X_0 , (z_k) has a convergent subsequence. W.l.o.g., we simply assume that $z_k \rightarrow z$, and by the infinite pigeonhole principle, we can further assume w.l.o.g. that $m_k = m$ is constant. Note first that since $z_k \in AF_k$ for all k , and since $z_k \rightarrow z$, we have $z \in F$ as F is explicitly closed. Note now that

$$|d(x_{m+1}, z) - d(x_{m+1}, z_k)|, |d(x_m, z) - d(x_m, z_k)| \leq d(z, z_k)$$

so that $d(x_{m+1}, z_k) \rightarrow d(x_{m+1}, z)$ and $d(x_m, z_k) \rightarrow d(x_m, z)$, uniformly. In particular, as G is continuous, let k_j be such that

$$|G(d(x_{m+1}, z)) - G(d(x_{m+1}, z_{k_j}))|, |G(d(x_m, z)) - G(d(x_m, z_{k_j}))| \leq \frac{1}{j+1}$$

for all $j \in \mathbb{N}$, uniformly on Ω . In particular, we have

$$G(d(x_{m+1}, z_{k_j})) \leq \frac{1}{j+1} + G(d(x_{m+1}, z)) \text{ and } G(d(x_m, z)) \leq \frac{1}{j+1} + G(d(x_m, z_{k_j})).$$

This thereby yields that if ω is such that

$$\mathbb{E}[G(d(x_{m+1}, z_{k_j})) \mid \mathcal{F}_m](\omega) > (1 + \chi_m(\omega))G(d(x_m(\omega), z_{k_j})) + \eta_m(\omega) + \frac{1}{r+1},$$

then we also have

$$-\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m](\omega) + (1 + \chi_m(\omega))G(d(x_m(\omega), z)) + \eta_m(\omega) + \frac{1}{r+1} < \frac{2 + \chi_m(\omega)}{j+1}.$$

Denoting the left random variable X_m , we hence have

$$\begin{aligned} & \mathbb{P} \left(X_m < \frac{2 + \chi_m}{j+1} \right) \\ & \geq \mathbb{P} \left(\mathbb{E}[G(d(x_{m+1}, z_{k_j})) \mid \mathcal{F}_m] > (1 + \chi_m)G(d(x_m, z_{k_j})) + \eta_m + \frac{1}{r+1} \right) \geq \lambda. \end{aligned}$$

By assumption, the sum $\sum_{n=0}^{\infty} \chi_n$ is almost surely bounded, and so also χ_m is almost surely bounded. So, let b_λ be an associated modulus of almost sure boundedness, that is $\mathbb{P}(\chi_m >$

$b_\lambda) < \lambda$ for all $\lambda > 0$. Hence, using the Fréchet inequality, we have

$$\mathbb{P}\left(X_m < \frac{2 + b_{\lambda/2}}{j+1}\right) \geq \mathbb{P}\left(X_m < \frac{2 + \chi_m}{j+1}\right) + \mathbb{P}(\chi_m \leq b_{\lambda/2}) - 1 \geq \lambda - \lambda/2 = \lambda/2.$$

However, we have

$$\lim_{j \rightarrow 0} \mathbb{P}\left(X_m < \frac{2 + b_{\lambda/2}}{j+1}\right) = \mathbb{P}(X_m \leq 0)$$

as $\mathbb{P}(X_m < \varepsilon)$ is right-continuous in ε . Now, however then

$$\mathbb{P}\left(\mathbb{E}[G(d(x_{m+1}, z)) \mid \mathcal{F}_m] \geq (1 + \chi_m)G(d(x_m, z)) + \eta_m + \frac{1}{r+1}\right) = \mathbb{P}(X_m \leq 0) \geq \lambda/2$$

and so the stochastic quasi-Fejér monotonicity is violated on a set with positive probability. Hence (x_n) is not stochastically quasi-Fejér monotone w.r.t. F , (AF_k) and \mathcal{F} , a contradiction.

By Theorem 4.21, we thus get that (x_n) is almost surely Cauchy. As X is complete and separable, we have that (x_n) almost surely converges to some random variable x . To show that $x \in F$ almost surely, note that (x_n) has the stochastic lim inf-property w.r.t. F and (AF_k) . Fix an ω such that $(x_n(\omega))$ satisfies

$$\forall k, N \in \mathbb{N} \exists n \geq N (x_n(\omega) \in AF_k)$$

and simultaneously $x_n(\omega) \rightarrow x(\omega)$. Using explicit closedness gives $x(\omega) \in F$. As the set of all such ω has measure 1, we get the result. \square

Remark 4.23. The above Theorem 4.22 can in fact be proved rather directly without appealing to Theorem 4.18 by following the folklore approach towards convergence of (stochastic) quasi-Fejér monotone sequences (see e.g. [17] as well as [50]), suitably adapted to (compact) metric spaces. Indeed, if (x_n) is stochastically quasi-Fejér monotone w.r.t. F , (AF_k) and \mathcal{F} , then by the Robbins-Siegmund theorem, we have that $G(d(x_n, z))$ converges a.s. for all $z \in F$, say on a set Ω_z with measure 1. By compactness, take (z_n) dense in F and denote by Ω' the set of measure 1 on which the stochastic lim inf-property holds for (x_n) w.r.t. F and (AF_k) . Let $\omega \in \widehat{\Omega} := \Omega' \cap \bigcap_{i \in \mathbb{N}} \Omega_{z_i}$. Then by $\omega \in \Omega'$, we have that there exists a subsequence x_{n_k} such that $x_{n_k}(\omega) \in AF_k$ for all $k \in \mathbb{N}$. By compactness, let $x^*(\omega)$ be a limit point of that sequence. Note that by explicit closedness of F , we have $x^*(\omega) \in F$. As (z_i) is dense in F , for any $\varepsilon > 0$ there exists an $i \in \mathbb{N}$ with $d(x^*(\omega), z_i) < \varepsilon$. As $\omega \in \Omega_{z_i}$, we have that $G(d(x_n(\omega), z_i))$ converges and by coercivity, $d(x_n(\omega), z_i)$ converges. In particular, $d(x^*(\omega), z_i)$ is a cluster point of the latter, so that given ε , there is some $i \in \mathbb{N}$ with

$$\lim_{n \rightarrow \infty} d(x_n(\omega), z_i) < \varepsilon.$$

Thus, we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} d(x_n(\omega), x^*(\omega)) &\leq \limsup_{n \rightarrow \infty} d(x_n(\omega), z_i) + d(x^*(\omega), z_i) \\ &\leq \lim_{n \rightarrow \infty} d(x_n(\omega), z_i) + \varepsilon \\ &\leq 2\varepsilon. \end{aligned}$$

Hence, as ε was arbitrary, we have $\lim_{n \rightarrow \infty} d(x_n(\omega), x^*(\omega)) = 0$.

The accomplishment of the above approach to Theorem 4.22 via Theorem 4.18 is thereby rather that it illustrates that Theorem 4.18 is a full finitization (in the sense of Tao, see [55]) of the result given in Theorem 4.22.

Remark 4.24. Similar to [35] (see Theorem 5.3 therein), if the assumption that F is explicitly closed is uniformly and quantitatively resolved via so-called moduli of uniform closedness, that is functions $\alpha_1, \alpha_2 : \mathbb{N} \rightarrow \mathbb{N}$ such that

$$\forall k \in \mathbb{N} \forall z, z' \in X \left(z \in AF_{\alpha_1(k)} \cap \overline{B}_{1/(\alpha_2(k)+1)}(z') \rightarrow z' \in AF_k \right),$$

then the previous quantitative convergence result given in Theorem 4.18 can be strengthened using these moduli to maintain that the process (x_n) contains approximate solutions along the region of metastability. Concretely, write $\alpha(k) := \max\{\alpha_1(k), \alpha_2(k)\}$ and $\widehat{k} := \max\{k, \alpha(k)\}$. Now, fix $\lambda > 0$, $k \in \mathbb{N}$ and $g \in \mathbb{N}^{\mathbb{N}}$. Following the proof of Theorem 4.18 (and therefore that of Lemma 4.16) with \widehat{k} instead of k , we get a set A with $\mathbb{P}(A^c) < \lambda$ such that, for any $\omega \in A$, there are numbers n_I, n_J with $n_I \leq \Delta(\lambda, \widehat{k}, g)$ and $x_{n_J}(\omega) \in AF_{\widehat{k}, \lambda}$ as well as

$$d(x_i(\omega), x_{n_J}(\omega)) \leq \frac{1}{2(\widehat{k} + 1)}$$

for all $i \in [n_I; n_I + g(n_I)]$. Using that $\widehat{k} \geq k$, we can conclude from this as before that

$$\forall i, j \in [n_I; n_I + g(n_I)] \left(d(x_i(\omega), x_j(\omega)) \leq \frac{1}{k + 1} \right).$$

Further however, note that

$$x_{n_J}(\omega) \in AF_{\widehat{k}, \lambda} \subseteq AF_{\widehat{k}, \alpha} \subseteq AF_{\widehat{k}} \subseteq AF_{\alpha_1(k)}$$

and so as $d(x_i(\omega), x_{n_J}(\omega)) \leq 1/(\alpha_2(k) + 1)$ for all $i \in [n_I; n_I + g(n_I)]$ since $\widehat{k} \geq \alpha_2(k)$, we get $x_i(\omega) \in AF_k$ for any $i \in [n_I; n_I + g(n_I)]$. Hence, we have shown that

$$\mathbb{P} \left(\forall n \leq \Delta(\lambda, \widehat{k}, g) \exists i, j \in [n; n + g(n)] \left(d(x_i, x_j) > \frac{1}{k + 1} \text{ or } x_i \notin AF_k \right) \right) \leq \mathbb{P}(A) < \lambda,$$

so that $x_i \in AF_k$ is maintained along the same region of metastability with high probability.

5. APPLICATION: COMMON FIXED POINT PROBLEMS

We end by outlining how the above results, both in their quantitative but also in their qualitative form, can be used to provide new convergence results for stochastic approximation procedures. We consider an example which incorporates many of the new notions we have introduced (while considerably simplifying others), namely a stochastic common fixed point problem and an associated randomized Krasnoselskii-Mann scheme, as e.g. recently investigated (in a broad context) over Hilbert spaces by Combettes and Madariaga [16].

Crucially however, our results are developed over the broad nonlinear context of separable Hadamard spaces, instead of just over Hilbert spaces, and our presentation is quite different, being tailored to the specifics setting of this paper in which we work over locally compact spaces, as opposed to assuming other regularity notions for guaranteeing convergence.

We begin by outlining our setup. As mentioned above, we will develop our results in the context of Hadamard spaces, that is complete geodesic metric spaces of nonpositive curvature in the sense of Alexandrov. These spaces, which uniformly cover examples such as Hilbert spaces, Hadamard manifolds (i.e. complete simply connected Riemannian manifolds of nonpositive sectional curvature) as well as \mathbb{R} -trees and beyond, frequently appear as one of the more canonical nonlinear generalizations of Hilbert spaces. We here only discuss the essential properties of these spaces as they are relevant to the present paper, and beyond that refer to [1, 7, 12] for further exposition, each of which have a different style and focus.

For this purpose of this paper, a metric space (X, d) is called (uniquely) geodesic if all two points are joined by a (unique) geodesic. In the uniquely geodesic case, given $x, y \in X$ and $\lambda \in [0, 1]$, we write $(1 - \lambda)x \oplus \lambda y$ for the unique point z on the geodesic connecting x and y satisfying $d(x, z) = \lambda d(x, y)$ and $d(y, z) = (1 - \lambda)d(x, y)$. A uniquely geodesic metric space (X, d) is a CAT(0) space if it satisfies

$$(CN) \quad d^2((1 - \lambda)x \oplus \lambda y, z) \leq (1 - \lambda)d^2(x, z) + \lambda d^2(y, z) - \lambda(1 - \lambda)d^2(x, y),$$

that is an extension of the so-called Bruhat-Tits CN-inequality to geodesics. A complete CAT(0) space is called a Hadamard space. We will here rely on another characterization of CAT(0) spaces due to Berg and Nikolaev [11], using their so-called quasi-inner product defined by

$$\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle := \frac{1}{2} (d^2(x, v) + d^2(y, u) - d^2(x, u) - d^2(y, v))$$

for all $x, y, u, v \in X$, where we write $\overrightarrow{xy}, \overrightarrow{uv}$ as a shorthand for pairs $(x, y), (u, v) \in X^2$. As shown in [11], already in a general metric space (X, d) , this represents the unique function $X^2 \times X^2 \rightarrow \mathbb{R}$ that satisfies

- (1) $\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle = d^2(x, y)$,
- (2) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle \overrightarrow{uv}, \overrightarrow{xy} \rangle$,
- (3) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = -\langle \overrightarrow{yx}, \overrightarrow{uv} \rangle$,
- (4) $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle + \langle \overrightarrow{xy}, \overrightarrow{vw} \rangle = \langle \overrightarrow{xy}, \overrightarrow{uw} \rangle$,

for all $x, y, u, v, w \in X$. It then follows from the results in [11] that a geodesic metric space (X, d) is a CAT(0) space if, and only if, the following metric version of the Cauchy-Schwartz inequality holds for all $x, y, u, v \in X$:

$$(CS) \quad \langle \overrightarrow{xy}, \overrightarrow{uv} \rangle \leq d(x, y)d(u, v).$$

Beyond the above, we in the following will only require a few other standard observations about CAT(0) spaces, which we collected in the following lemma.

Lemma 5.1. *Let (X, d) be a CAT(0) space. Then*

$$(*)_1 \quad d^2(x, z) = d^2(x, y) + d^2(y, z) + 2 \langle \overrightarrow{xy}, \overrightarrow{yz} \rangle$$

for all $x, y, z \in X$, as well as

$$(*)_2 \quad d((1 - \lambda)x \oplus \lambda y, z) \leq (1 - \lambda)d(x, z) + \lambda d(y, z)$$

for all $x, y, z \in X$ and $\lambda \in [0, 1]$.

The first of the above properties is the result of a direct calculation, the second immediately follows from the joint convexity of the metric in CAT(0) (see e.g. in Proposition 1.1.5 in [7]). We hence omit the proofs.

5.1. Common fixed points on separable Hadamard spaces. For the remainder of this section, we assume that (X, d) is some proper Hadamard space (which is hence separable), and let $\mathcal{B}(X)$ denote the usual Borel σ -algebra on X . Beyond these, we assume an ambient probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Fixing some other measure space (V, \mathcal{V}) , we let $(T_v)_{v \in V}$ be a family of nonexpansive mappings on X i.e.

$$d(T_v x, T_v y) \leq d(x, y)$$

for all $x, y \in X$ and $v \in V$. We assume furthermore that the mapping $V \times X \rightarrow X$ given by $(v, x) \mapsto T_v x$ is $\mathcal{V} \otimes \mathcal{B}(X)/\mathcal{B}(X)$ -measurable. We now let $v : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (V, \mathcal{V})$ be a V -valued random variable over our ambient probability space, and consider the following problem:

$$\text{find some element of } F := \{z \in X \mid T_v z = z \text{ } \mathbb{P}\text{-a.s.}\}.$$

Throughout, we assume that $F \neq \emptyset$. Crucially, we can express F equivalently by

$$F := \{z \in X \mid \mathbb{E}[d^2(T_v z, z)] = 0\},$$

which motivates a sequence of approximating sets (AF_k) defined by

$$AF_k := \left\{ z \in X_0 \mid \mathbb{E}[d^2(T_v z, z)] \leq \frac{1}{k+1} \right\},$$

where we already incorporate a closed set $X_0 \subseteq X$, which will later be concretely instantiated in the context of our method.

Indeed, it is then immediate that (AF_k) approximate F in the sense that $AF_{k+1} \subseteq AF_k$ for all $k \in \mathbb{N}$ and that $\bigcap_{k \in \mathbb{N}} AF_k = F$. Further, also the following result is rather immediate:

Lemma 5.2. *For any $k \in \mathbb{N}$, AF_k is closed. In particular, F is closed.*

Proof. Fix $k \in \mathbb{N}$, and let $(z_n) \subseteq AF_k$ with $z_n \rightarrow z$. Note that almost surely, we then have $d(T_v z, z) \leq d(T_v z_n, z_n) + 2d(z_n, z)$ and so

$$d^2(T_v z, z) \leq d^2(T_v z_n, z_n) + 4d(z_n, z)d(T_v z_n, z_n) + 4d^2(z_n, z)$$

for all $n \in \mathbb{N}$. Now, note that $\mathbb{E}[d^2(T_v z_n, z_n)] \leq \frac{1}{k+1}$ and so $\mathbb{E}[d(T_v z_n, z_n)] \leq \sqrt{\frac{1}{k+1}}$. Applying expectations, we get

$$\begin{aligned} \mathbb{E}[d^2(T_v z, z)] &\leq \mathbb{E}[d^2(T_v z_n, z_n)] + 4d(z_n, z)\mathbb{E}[d(T_v z_n, z_n)] + 4d^2(z_n, z) \\ &\leq \frac{1}{k+1} + 4d(z_n, z)\sqrt{\frac{1}{k+1}} + 4d^2(z_n, z). \end{aligned}$$

Taking limits yields $\mathbb{E}[d^2(T_v z, z)] \leq \frac{1}{k+1}$ and so $z \in AF_k$. \square

In particular, we thereby get that all AF_k and also F is measurable. Further, we also get that F is explicitly closed w.r.t. (AF_k) in the following particularly uniform way (recall Remark 4.24):

Lemma 5.3. *Fix $k \in \mathbb{N}$. For any $z, z' \in X$, if $z \in AF_{12k+11} \cap \overline{B}_{1/12(k+1)}(z')$, then $z' \in AF_k$.*

Proof. Almost surely, we have $d(T_v z', z') \leq d(T_v z, z) + 2d(z', z)$ as before. Also as before, we have $\mathbb{E}[d(T_v z, z)] \leq \sqrt{\frac{1}{12(k+1)}}$. Squaring and taking expectations yields

$$\begin{aligned} \mathbb{E}[d^2(T_v z', z')] &\leq \mathbb{E}[d^2(T_v z, z)] + 4d(z', z)\mathbb{E}[d(T_v z, z)] + 4d^2(z', z) \\ &\leq \frac{1}{12(k+1)} + \frac{4}{12(k+1)} + \frac{4}{12(k+1)} \leq \frac{1}{k+1} \end{aligned}$$

which gives $z' \in AF_k$. \square

Remark 5.4. We can define a sequence of sets

$$AF'_k := \left\{ z \in X_0 \mid \mathbb{P} \left(d^2(T_v z, z) \geq \frac{1}{k+1} \right) \leq \frac{1}{k+1} \right\}.$$

which similarly approximate F , and which simultaneously stays closer in character to the phrasing of F as an almost-sure equation. Indeed, we have

$$\mathbb{P} \left(d^2(T_v z, z) \geq \frac{1}{k+1} \right) \leq (k+1)\mathbb{E}[d^2(T_v z, z)]$$

using Markov's inequality, so that $AF_{k^2+2k} \subseteq AF'_k$.

Conversely, assume X_0 has bounded diameter and let $c \in \mathbb{N}^*$ be a bound on this diameter. Using $F \neq \emptyset$, let $p \in F$. Note that then $d(T_v z, z) \leq d(T_v p, p) + 2d(p, z) \leq 2c$ for all $z \in X_0$. We then have $AF'_{\rho(k)} \subseteq AF_k$, with $\rho(k) := \max\{2k + 1, 4c^2(2k + 2) - 1\}$. Indeed, assume that $z \in AF'_{\rho(k)}$, that is

$$\mathbb{P}\left(d^2(T_v z, z) \geq \frac{1}{\rho(k) + 1}\right) \leq \frac{1}{\rho(k) + 1}.$$

Denote the inner set by A . Then we have $\mathbb{E}[d^2(T_v z, z)\mathbf{1}_A] \leq 4c^2\mathbb{P}(A) \leq \frac{1}{2(k+1)}$, as well as $\mathbb{E}[d(T_v z, z)\mathbf{1}_{A^c}] \leq \frac{1}{2(k+1)}$, so that combined we have

$$\mathbb{E}[d(T_v z, z)] = \mathbb{E}[d(T_v z, z)\mathbf{1}_A] + \mathbb{E}[d(T_v z, z)\mathbf{1}_{A^c}] \leq \frac{1}{k+1},$$

that is $z \in AF_k$.

However, the above formulation of (AF_k) using the mean proves to be considerably more productive for the method studied in the following section, so that we have opted for that specific representation.

5.2. A Krasnoselskii-Mann scheme for approximating common fixed points. We now consider a simple Krasnoselskii-Mann scheme for approximating elements of F . Let (x_n) be a sequence of X -valued random variables defined by

$$(sKM) \quad x_{n+1} := (1 - \lambda_n)x_n \oplus \lambda_n T_{v_n} x_n$$

for some $x_0 \in X$, where $(\lambda_n) \subseteq (0, 1]$ and (v_n) is a sequence of independent V -valued random variables distributed as v . In particular, v_n is independent of $\mathcal{F}_n := \sigma(x_0, \dots, x_n)$ for all $n \in \mathbb{N}$. As geodesics are continuous in their endpoints (see e.g. Lemma 1.2.2 in [7]), it immediately follows by induction that each x_n is measurable.

As before, we assume that $F \neq \emptyset$ and now fix some $p \in F$, choosing $c \geq 1$ to satisfy $d(x_0, p) \leq c$.

Lemma 5.5. *For any $z \in X_0$ and $n \in \mathbb{N}$ we have $d(x_n, p) \leq c$ almost surely.*

Proof. For $n \in \mathbb{N}$, since v_n is distributed as v and $p \in F$, we have $d(T_{v_n} p, p) = 0$ almost surely. Now, using $(*)_2$, it follows that

$$\begin{aligned} d(x_{n+1}, p) &\leq (1 - \lambda_n)d(x_n, p) + \lambda_n d(T_{v_n} x_n, p) \\ &\leq (1 - \lambda_n)d(x_n, p) + \lambda_n (d(T_{v_n} x_n, T_{v_n} p) + d(T_{v_n} p, p)) \\ &\leq d(x_n, p) \end{aligned}$$

so that $d(x_n, p) \leq d(x_0, p) \leq c$ for any $n \in \mathbb{N}$, almost surely. \square

In the following, we will thereby work over the associated space $X_0 = \overline{B}_c(p)$, which is compact as X is proper. In particular, note that we have $d(x_n, z) \leq 2c$ for any $z \in X_0$ and $n \in \mathbb{N}$.

Remark 5.6. Note that we by Lemma 5.5 in particular have that $\mu(\varepsilon) := \varepsilon/4c^2$ is a modulus of absolute continuity for any $d^2(x_n, z)$, since

$$\mathbb{E}[d^2(x_n, z)\mathbf{1}_A] \leq 4c^2\mathbb{P}(A) \leq \varepsilon$$

for any $A \in \mathcal{F}$ with $\mathbb{P}(A) < \varepsilon/4c^2$. Further, we can construct a modulus of almost sure boundedness for $(d(x_n, p))$ by setting $b_\lambda := c$, as we have

$$\mathbb{P}(\exists n \in \mathbb{N} (d(x_n, p) > c)) = 0 < \lambda$$

for any $\lambda > 0$.

Lemma 5.7. *For any $n \in \mathbb{N}$ and $z \in X$ we have*

$$\begin{aligned} & \mathbb{E} [d^2(x_{n+1}, z) \mid \mathcal{F}_n] \\ & \leq (1 - \lambda_n)d^2(x_n, z) + \lambda_n \mathbb{E} [d^2(T_{v_n}x_n, z) \mid \mathcal{F}_n] - \lambda_n(1 - \lambda_n)\mathbb{E} [d^2(x_n, T_{v_n}x_n) \mid \mathcal{F}_n]. \end{aligned}$$

Proof. By (CN) we have

$$d^2(x_{n+1}, z) \leq (1 - \lambda_n)d^2(x_n, z) + \lambda_n d^2(T_{v_n}x_n, z) - \lambda_n(1 - \lambda_n)d^2(x_n, T_{v_n}x_n)$$

and the result then follows by taking conditional expectations and using that $d(x_n, z)$ is \mathcal{F}_n -measurable. \square

Lemma 5.8. *The scheme (x_n) satisfies*

$$\forall r, n \in \mathbb{N} \forall z \in AF_{64c^2(r+1)} \forall m \leq n \left(\mathbb{E} [d^2(x_{m+1}, z) \mid \mathcal{F}_m] \leq d^2(x_m, z) + \frac{1}{r+1} \right).$$

and is therefore uniformly stochastic quasi-Fejér monotone w.r.t. F , (AF_k) and (\mathcal{F}_n) , and more specifically for $G(a) := a^2$, $\chi_m := \eta_m := 0$, and $\zeta(\lambda, r, n) := 64c^2(r+1)^2$.

Proof. For any $m \in \mathbb{N}$ and $z \in X_0$, we first observe that we have

$$\begin{aligned} d^2(T_{v_m}x_m, z) & \leq d^2(T_{v_m}x_m, T_{v_m}z) + d^2(T_{v_m}z, z) + 2d(T_{v_m}x_m, T_{v_m}z)d(T_{v_m}z, z) \\ & \leq d^2(x_m, z) + d^2(T_{v_m}z, z) + 2d(x_m, z)d(T_{v_m}z, z) \\ & \leq d^2(x_m, z) + d^2(T_{v_m}z, z) + 4cd(T_{v_m}z, z), \end{aligned}$$

where the last line follows from Lemma 5.5. Therefore

$$\begin{aligned} \mathbb{E} [d^2(T_{v_m}x_m, z) \mid \mathcal{F}_m] & \leq d^2(x_m, z) + \mathbb{E} [d^2(T_{v_m}z, z) \mid \mathcal{F}_m] + 4c\mathbb{E} [d(T_{v_m}z, z) \mid \mathcal{F}_m] \\ & = d^2(x_m, z) + \mathbb{E} [d^2(T_vz, z)] + 4c\mathbb{E} [d(T_vz, z)] \end{aligned}$$

for the second inequality using that v_m is independent of \mathcal{F}_m and distributed as v . By Lemma 5.7 we therefore have

$$\begin{aligned} \mathbb{E} [d^2(x_{m+1}, z) \mid \mathcal{F}_m] & \leq (1 - \lambda_m)d(x_m, z) + \lambda_m \mathbb{E} [d^2(T_{v_m}x_m, z) \mid \mathcal{F}_m] \\ & \leq d^2(x_m, z) + \lambda_m (\mathbb{E} [d^2(T_vz, z)] + 4c\mathbb{E} [d(T_vz, z)]) \\ & \leq d^2(x_m, z) + \mathbb{E} [d^2(T_vz, z)] + 4c\mathbb{E} [d(T_vz, z)] \end{aligned}$$

for any $m \in \mathbb{N}$ and $z \in X_0$, using $\lambda_m \leq 1$ in the last line. Now fix $r, n \in \mathbb{N}$, $z \in AF_{64c^2(r+1)^2}$ and $m \leq n$. Using Jensen's inequality, it follows that

$$\mathbb{E} [d^2(T_vz, z)] \leq \frac{1}{2(r+1)} \text{ and } \mathbb{E} [d(T_vz, z)] \leq \frac{1}{8c(r+1)}$$

from which the result follows. \square

Remark 5.9. Our scheme is stochastically quasi-Fejér monotone in an extremely simple way, where in particular the error sequences (χ_m) , (η_m) disappear and the modulus $\zeta(\lambda, r, n)$ depends only on r , and indeed we deliberately chose an example through which our main results could be illustrated in a simplified setting. However, even in this simple setting, there are small extensions that would lead to more elaborate versions of the corresponding quasi-Fejér property. For example, one could study an extension of our scheme with noise terms e_n :

$$x_{n+1} := (1 - \lambda_n)x_n \oplus \lambda_n y_n \quad \text{where} \quad d(y_n, T_{v_n}x_n) \leq e_n \quad \mathbb{P}\text{-a.s.}$$

A simple instance of this scheme in a Hilbert space and for non-random mappings, given via

$$x_{n+1} := x_n + \lambda_n(Tx_n + e_n - x_n)$$

was considered in [17] in the presence of compactness, giving rise to a stochastic quasi-Fejér property in which the noise terms are now incorporated into the sequence (η_m) . Another form of generality would be obtained by weakening the i.i.d. assumption of the (v_n) to

$$\forall n \in \mathbb{N} \forall z \in X \left(z \in F \rightarrow \mathbb{E} \left[d^2(T_{v_n} z, z) \mid \mathcal{F}_n \right] = 0 \right),$$

and then defining the (AF_k) implicitly through a quantitative rendering of that property, namely an abstract modulus ρ satisfying

$$\forall n, k \in \mathbb{N} \forall z \in X \left(z \in AF_{\rho(n,k)} \rightarrow \mathbb{E} \left[d^2(T_{v_n} z, z) \mid \mathcal{F}_n \right] \leq \frac{1}{k+1} \right).$$

A simple adaptation of Lemma 5.8 would then establish uniform stochastic quasi-Fejér monotonicity of (x_n) where now $\zeta(\lambda, r, n) := \rho(n, 64c^2(r+1)^2)$ is also dependent on n . However, we leave a proper study of various generalisations of the scheme discussed here, along with further concrete examples of our framework, to future work.

Having shown that the scheme (x_n) is stochastically quasi-Fejér monotone, we now show how to define an explicit lim inf-modulus Φ , under the additional (standard) assumption that the step-sizes (λ_n) satisfy $\sum_{n=0}^{\infty} \lambda_n(1-\lambda_n) = \infty$. Quantitatively, this assumption is resolved using a rate of divergence $\theta : \mathbb{N} \times (0, \infty) \rightarrow \mathbb{N}$, i.e.

$$\forall N \in \mathbb{N} \forall b > 0 \left(\sum_{n=N}^{\theta(N,b)} \lambda_n(1-\lambda_n) \geq b \right).$$

We begin by identifying a simpler lim inf property, widely seen in the convergence analysis of Krasnoselskii-Mann-type schemes:

Lemma 5.10. *Define $\Psi : (0, \infty) \times \mathbb{N} \rightarrow \mathbb{N}$ by*

$$\Psi(\mu, N) := \theta \left(N, \frac{c^2 + 1}{\mu} \right).$$

Then we have

$$\forall \mu > 0 \forall N \in \mathbb{N} \exists n \in [N; \Psi(\mu, N)] \left(\mathbb{E} \left[\int_{\Omega} d^2(T_{v(\omega)} x_n, x_n) \mathbb{P}(d\omega) \right] < \mu \right).$$

Proof. For any $n \in \mathbb{N}$, by CN we have

$$\begin{aligned} d^2(x_{n+1}, p) &\leq (1-\lambda_n)d^2(x_n, p) + \lambda_n d^2(T_{v_n} x_n, p) - \lambda_n(1-\lambda_n)d^2(T_{v_n} x_n, x_n) \\ &\leq d^2(x_n, p) - \lambda_n(1-\lambda_n)d^2(T_{v_n} x_n, x_n) \end{aligned}$$

where here we use that $d(T_{v_n} p, p) = 0$ and hence $d(T_{v_n} x_n, p) \leq d(x_n, p)$ almost surely. But then for any $N \in \mathbb{N}$ we have

$$\sum_{n=N}^k \lambda_n(1-\lambda_n)d^2(T_{v_n} x_n, x_n) \leq d^2(x_N, p) - d^2(x_{k+1}, p) \leq d^2(x_N, p) \leq c^2$$

and therefore, if $\mathbb{E} [d^2(T_{v_n} x_n, x_n)] \geq \mu$ for all $n \in [N; \Psi(\mu, N)]$, then

$$c^2 + 1 \leq \mu \sum_{n=N}^{\theta(N, (c^2+1)/\mu)} \lambda_n(1-\lambda_n) \leq \sum_{n=N}^{\theta(N, (c^2+1)/\mu)} \lambda_n(1-\lambda_n) \mathbb{E} [d^2(T_{v_n} x_n, x_n)] \leq c^2,$$

a contradiction. The result follows as we have

$$\mathbb{E} [d^2(T_{v_n} x_n, x_n)] = \mathbb{E} \left[\int_{\Omega} d^2(T_{v(\omega)} x_n, x_n) \mathbb{P}(d\omega) \right]$$

from Fubini's theorem, and since v_n is independent of x_n and distributed as v . \square

Lemma 5.11. *Let Ψ be defined as in Lemma 5.10, and set*

$$\Phi(\lambda, k, N) := \Psi\left(\frac{\lambda}{k+1}, N\right) = \theta\left(N, \frac{(c^2+1)(k+1)}{\lambda}\right).$$

Then Φ is a stochastic lim inf-modulus w.r.t. F and (AF_k) for (x_n) i.e.

$$\forall \lambda > 0 \forall k, N \in \mathbb{N} (\mathbb{P}(\forall n \in [N; \Phi(\lambda, k, N)](x_n \notin AF_k)) < \lambda).$$

Proof. Fix $\lambda > 0$ and $k, N \in \mathbb{N}$, and define $\mu := \lambda/(k+1)$. By Lemma 5.10, there exists $n \in [N; \Phi(\lambda, k, N)]$ such that

$$\mathbb{E}\left[\int_{\Omega} d^2(T_{v(\omega)}x_n, x_n) \mathbb{P}(d\omega)\right] < \mu.$$

We then have

$$\mathbb{P}(x_n \notin AF_k) = \mathbb{P}\left(\int_{\Omega} d^2(T_{v(\omega)}x_n, x_n) \mathbb{P}(d\omega) > \frac{1}{k+1}\right) \leq (k+1)\mathbb{E}\left[\int_{\Omega} d^2(T_{v(\omega)}x_n, x_n) \mathbb{P}(d\omega)\right]$$

by Markov's inequality, and so $\mathbb{P}(x_n \notin AF_k) < \lambda$. \square

We now have everything we need to instantiate the main theorem of the paper, by which we can derive the following (quantitative) convergence theorem for the above stochastic Krasnoselskii-Mann-type scheme.

Theorem 5.12. *Let (X, d) be a proper Hadamard space, (V, \mathcal{V}) a measure space, and $(T_v)_{v \in V}$ a family of nonexpansive mappings on X such that $(v, x) \mapsto T_v x$ is measurable.*

Let v be a V -valued random variable, and define

$$F := \{z \in X \mid T_v z = z \text{ } \mathbb{P}\text{-a.s.}\}.$$

As before, we assume that $F \neq \emptyset$ and fix some $p \in F$ as well as a $c \geq 1$ with $d(x_0, p) \leq c$. Set $X_0 := \overline{B}_c(p)$. Define the scheme (x_n) by

$$x_{n+1} := (1 - \lambda_n)x_n \oplus \lambda_n T_{v_n} x_n$$

for some $x_0 \in X$, $(\lambda_n) \subset (0, 1]$ with $\sum_{n=0}^{\infty} \lambda_n(1 - \lambda_n) = \infty$, and (v_n) a sequence of independent V -valued random variables distributed as v .

Then $x_n \rightarrow x$ almost surely, where $x \in F$ almost surely. Furthermore, if $\gamma(k)$ is a modulus of total boundedness for X_0 , $\theta(N, b)$ is a rate of divergence for $\sum_{n=0}^{\infty} \lambda_n(1 - \lambda_n)$, then a rate of metastability for (x_n) is explicitly definable in those data.

Proof. The first part follows from Theorem 4.22 for $G(a) := a^2$ and (AF_k) as defined in Section 5.1, where in particular (x_n) is stochastically quasi-Fejér monotone by Lemma 5.8 (by which it also satisfies the stronger property of being uniformly stochastic quasi-Fejér monotone), has the stochastic lim inf property by Lemma 5.11, and is explicitly closed by Lemma 5.3. For the second part, we appeal to Theorem 4.18, which in addition to the modulus of uniform boundedness $\gamma(k)$, has its assumptions satisfied as follows:

- the modulus of continuity and inverse continuity for $G(a) := a^2$ are given by $\iota_b(k) := 2b(k+1)$ and $\nu(k) := (k+1)^2$, respectively;
- we set $o := p$ and then c is by definition a bound relative to p for X_0 ;
- $\eta_n := \chi_n := 0$ so we can set $f(\lambda) := 1$ and $h(\lambda) := 2$;
- $\zeta(\lambda, r, n) := 64c^2(r+1)^2$ as defined in Lemma 5.8;
- $\Phi(\lambda, k, N) := \theta(N, (2c+1)(k+1)/\lambda)$ as defined in Lemma 5.11;
- since $d(x_0, p) + c \leq 2c$ we have $\mathbb{E}[(d(x_0, p) + c)^2] < b_0$ for $b_0 := 4c^2 + 1$.

Recall also Remark 5.6 for definitions of moduli of absolute continuity and of almost sure boundedness for the respective sequence in terms of c .

Therefore, the rate of metastability as given in Theorem 4.18 depends only on $\gamma(k)$, c and $\theta(N, b)$. \square

Acknowledgments. The authors want to thank Pedro Pinto for helpful discussions on the topic of this paper.

Funding. The first author was partially supported by the EPSRC Centre for Doctoral Training in Digital Entertainment EP/L016540/1, and the third author was partially supported by the EPSRC grant EP/W035847/1. The second author was partially supported by the Deutsche Forschungsgemeinschaft Project DFG KO 1737/6-2.

REFERENCES

- [1] S. Alexander, V. Kapovitch, and A. Petrunin. *Alexandrov Geometry: Foundations*, volume 236 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2024.
- [2] A. Arakcheev and H.H. Bauschke. Fejér and Fejér* Monotonicity: New Results and Limiting Examples, 2025. Preprint, available at <https://arxiv.org/abs/2512.17039>.
- [3] A. Arakcheev and H.H. Bauschke. On Opial’s Lemma, 2025. Preprint, available at <https://arxiv.org/abs/2503.22004>.
- [4] J. Avigad, P. Gerhardy, and H. Towsner. Local stability of ergodic averages. *Transactions of the American Mathematical Society*, 362(1):261–288, 2010.
- [5] H.H. Bauschke, J.M. Borwein, and P.L. Borwein. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.
- [6] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer Cham, 2nd edition, 2017.
- [7] M. Bačák. *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. Walter de Gruyter GmbH, Berlin/Boston, 2014.
- [8] M. Bačák, I. Searston, and B. Sims. Alternating projections in CAT(0) spaces. *Journal of Mathematical Analysis and Applications*, 385:599–607, 2012.
- [9] R. Behling, Y. Bello-Cruz, A. Iusem, D. Liu, and L.R. Santos. A finitely convergent circumcenter method for the convex feasibility problem. *SIAM Journal on Optimization*, 34:2535–2556, 2024.
- [10] R. Behling, Y. Bello-Cruz, A.N. Iusem, A. Alves Ribeiro, and L.-R. Santos. Fejér* monotonicity in optimization algorithms, 2024. Preprint, available at <https://arxiv.org/abs/2410.08331>.
- [11] I.D. Berg and I.G. Nikolaev. Quasilinearization and curvature of Aleksandrov spaces. *Geometriae Dedicata*, 133:195–218, 2008.
- [12] M.R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg, 1999.
- [13] A.A. Chashka. Fluctuations in martingales. *Uspekhi Matematicheskikh Nauk*, 49(2):179–180, 1994. English translation in *Russian Math. Surveys* 49:2 (1994).
- [14] P.L. Combettes. Quasi-Fejérian Analysis of Some Optimization Algorithms. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, volume 8 of *Studies in Computational Mathematics*, pages 115–152. North-Holland, Amsterdam, 2001.
- [15] P.L. Combettes. Fejér monotonicity in convex optimization. In C.A. Floudas and P.M. Pardalos, editors, *Encyclopedia of Optimization*, pages 1016–1024. Springer, Boston, MA, 2009.
- [16] P.L. Combettes and J.I. Madariaga. A Geometric Framework for Stochastic Iterations, 2025. Preprint, available at <https://arxiv.org/abs/2504.02761>.
- [17] P.L. Combettes and J.-C. Pesquet. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. *SIAM Journal on Optimization*, 25(2):1221–1248, 2015.
- [18] P.L. Combettes and J.-C. Pesquet. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: mean-square and linear convergence. *Mathematical Programming*, 174(1):433–451, 2019.
- [19] P.L. Combettes and B.C. Vũ. Variable metric quasi-Fejér monotonicity. *Nonlinear Analysis: Theory, Methods & Applications*, 78:17–31, 2013.
- [20] J.L. Doob. *Stochastic Processes*. Wiley, N.Y., 1953.

- [21] J.L. Doob. Notes on Martingale Theory. In *Berkeley Symposium on Mathematical Statistics and Probability*, pages 95–102. University of California Press, 1961.
- [22] I.I. Eremin. Methods of Fejér approximations in convex programming. *Mathematical notes of the Academy of Sciences of the USSR*, 3:139–149, 1968.
- [23] I.I. Eremin. On the speed of convergence in the method of Fejér approximations. *Mathematical notes of the Academy of Sciences of the USSR*, 4:522–527, 1968.
- [24] Y.M. Ermol'ev. On the method of generalized stochastic gradients and quasi-Fejér sequences. *Cybernetics*, 5:208–220, 1969.
- [25] Y.M. Ermol'ev. On convergence of random quasi-Fejér sequences. *Cybernetics*, 7:655–656, 1971.
- [26] Y.M. Ermol'ev and A.D. Tuniev. Random Fejér and quasi-Fejér sequences. *Theory of Optimal Solutions – Akademiya Nauk Ukrainської SSR Kiev*, 2:76–83, 1968. in Russian; English translation in Amer. Math. Soc. Select. Translat. Math. Statist. Probab., 13 (1973), pp. 143–148.
- [27] L. Fejér. Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen. *Mathematische Annalen*, 85:41–48, 1922.
- [28] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame Journal of Formal Logic*, 49:431–446, 2008.
- [29] R.L. Jones, R. Kaufman, J.M. Rosenblatt, and M. Wierdl. Oscillation in ergodic theory. *Ergodic Theory and Dynamical Systems*, 18(4):889–935, 1998.
- [30] A.G. Kachurovskii. The rate of convergence in ergodic theorems. *Russian Mathematical Surveys*, 51(4):653–703, 1996.
- [31] A. Klenke. *Probability Theory*. Springer, 3rd edition, 2020.
- [32] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.
- [33] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proceedings of ICM 2018*, volume 2, pages 61–82. World Scientific, Singapore, 2019.
- [34] U. Kohlenbach. Local formalizations in nonlinear analysis and related areas and proof-theoretic tameness. In P. Weingartner and H.-P. Leeb, editors, *Kreisel's Interests. On the Foundations of Logic and Mathematics*, volume 41 of *Tributes*, pages 45–61. College Publications, London, 2020.
- [35] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20(2), 2018. 1750015, 42pp.
- [36] U. Kohlenbach, G. Lopez-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232:261–297, 2019.
- [37] U. Kohlenbach and P. Oliva. Proof Mining: A Systematic Way of Analysing Proofs in Mathematics. *Proceedings of the Steklov Institute of Mathematics*, 242:136–164, 2003.
- [38] U. Kohlenbach and P. Pinto. Fejér Monotone Sequences Revisited. *Journal of Convex Analysis*, 2025. Preprint available at <https://arxiv.org/abs/2310.06528>.
- [39] U. Kohlenbach and P. Safarik. Fluctuations, effective learnability and metastability in analysis. *Annals of Pure and Applied Logic*, 165:266–304, 2014.
- [40] M. Métivier. *Semimartingales*, volume 2 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 1982.
- [41] T. Motzkin and I. Schoenberg. The Relaxation Method for Linear Inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.
- [42] M. Neri, P. Oliva, and N. Pischke. A systematic way of analysing proofs in probability theory, 2026. Manuscript in preparation.
- [43] M. Neri and N. Pischke. Proof mining and probability theory. *Forum of Mathematics, Sigma*, 13, 2025. e187, 47pp.
- [44] M. Neri, N. Pischke, and T. Powell. On the asymptotic behaviour of stochastic processes, with applications to supermartingale convergence, Dvoretzky's approximation theorem, and stochastic quasi-Fejér monotonicity, 2025. Preprint, available at <https://arxiv.org/abs/2504.12922>.
- [45] M. Neri and T. Powell. On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales. *Transactions of the American Mathematical Society, Series B*, 12:974–1019, 2025.
- [46] M. Neri and T. Powell. A quantitative Robbins-Siegmund theorem. *Annals of Applied Probability*, 36(1):636–651, 2026.
- [47] Morenikeji Neri, Nicholas Pischke, and Thomas Powell. Generalized learnability of stochastic principles. In *Proceedings of Computability in Europe (CiE'25)*, volume 15764 of *LNCS*, pages 333–348, 2025.
- [48] E. Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11(4:20), 2015. 44pp.

- [49] N. Pischke. Generalized Fejér monotone sequences and their finitary content. *Optimization*, 74(14):3771–3838, 2025.
- [50] N. Pischke. On Busemann subgradient methods for stochastic minimization in Hadamard spaces, 2026. Preprint, available at <https://arxiv.org/abs/2602.08127>.
- [51] N. Pischke and T. Powell. Asymptotic regularity of a generalised stochastic Halpern scheme, 2024. Preprint, available at <https://arxiv.org/abs/2411.04845>.
- [52] H. Robbins and D. Siegmund. A convergence theorem for non-negative almost supermartingales and some applications. In *Optimizing methods in statistics*, pages 233–257. Elsevier, 1971.
- [53] E. Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *The Journal of Symbolic Logic*, 14:145–208, 1949.
- [54] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28(2):657–688, 2008.
- [55] T. Tao. *Soft Analysis, Hard Analysis, and the Finite Convergence Principle*, pages 17–29. American Mathematical Society, Providence, RI, 2008.
- [56] J. Ville. *Étude Critique de la Notion de Collectif*. PhD thesis, École Polytechnique, 1939.
- [57] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.